Measuring Cross-Modal Interactions in Multimodal Models

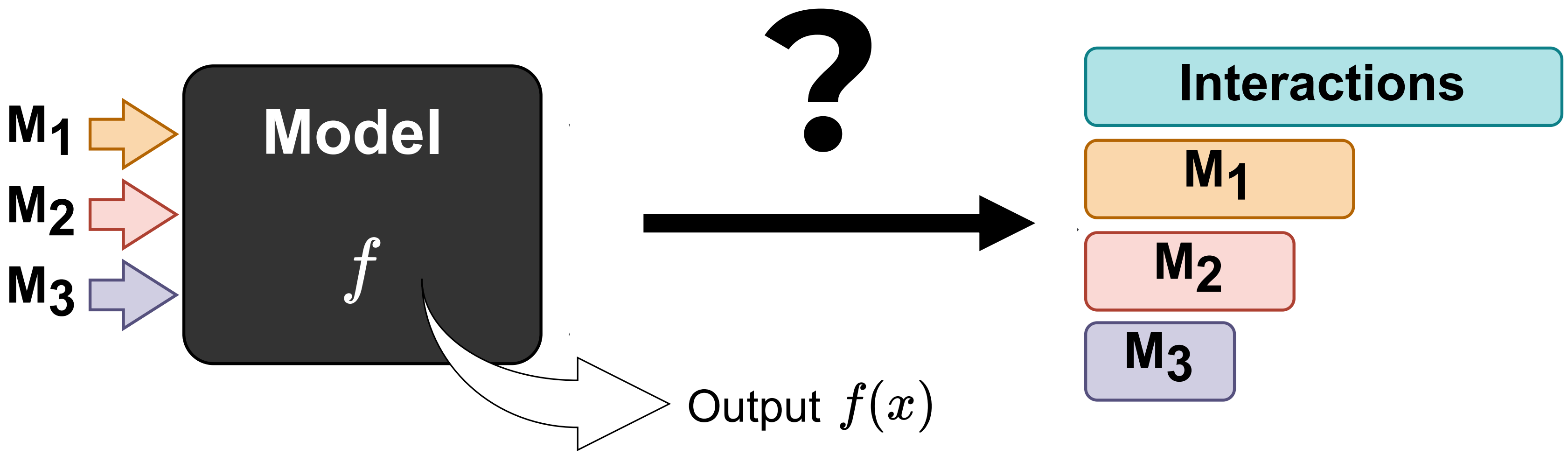
Laura Wenderoth[1] , Konstantin Hemker[1], Nikola Simidjievski[2,1], Mateja Jamnik[1]

[1]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK

[2]PBCI, Department of Oncology, University of Cambridge, Cambridge, United Kingdom

{lw457, kh701, ns779, mj201}@cam.ac.uk

Motivation

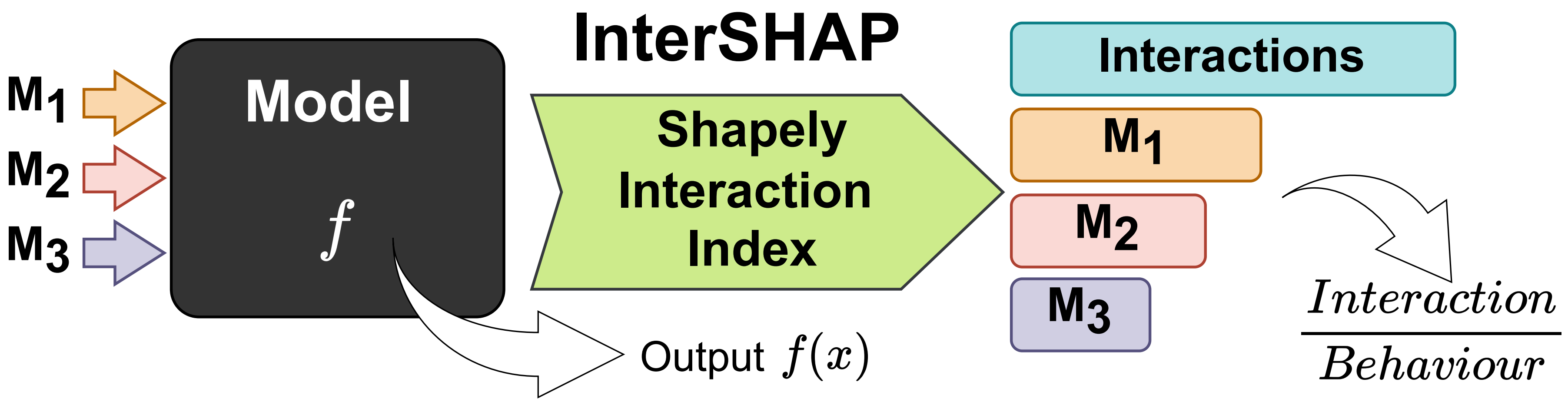


Limitations of Previous Work

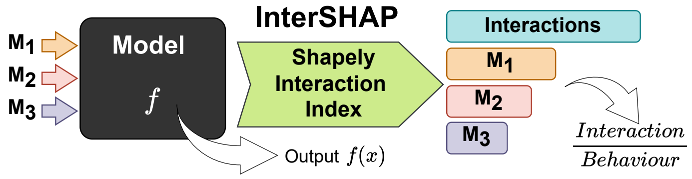
InterSHAP overcomes the limitations of other cross-modal interaction scores: it is unsupervised, performance agnostic, applicable to more than two modalities, and allows for dataset- (global) and sample-level (local) explainability

Score	Modalities > 2	Local	Unsupervised	Performance Agnostic
PID[1]	✗	✗	✓	✓
EMAP[2]	✗	✗	✗	✗
SHAPE[3]	✓	✗	✗	✗
InterSHAP (Ours)	✓	✓	✓	✓

InterSHAP


M1
M2
M3
Model
f
Output f(x)
InterSHAP
Shapely
Interaction
Index
Interactions
M1
M2
M3
Interaction
Behaviour

InterSHAP



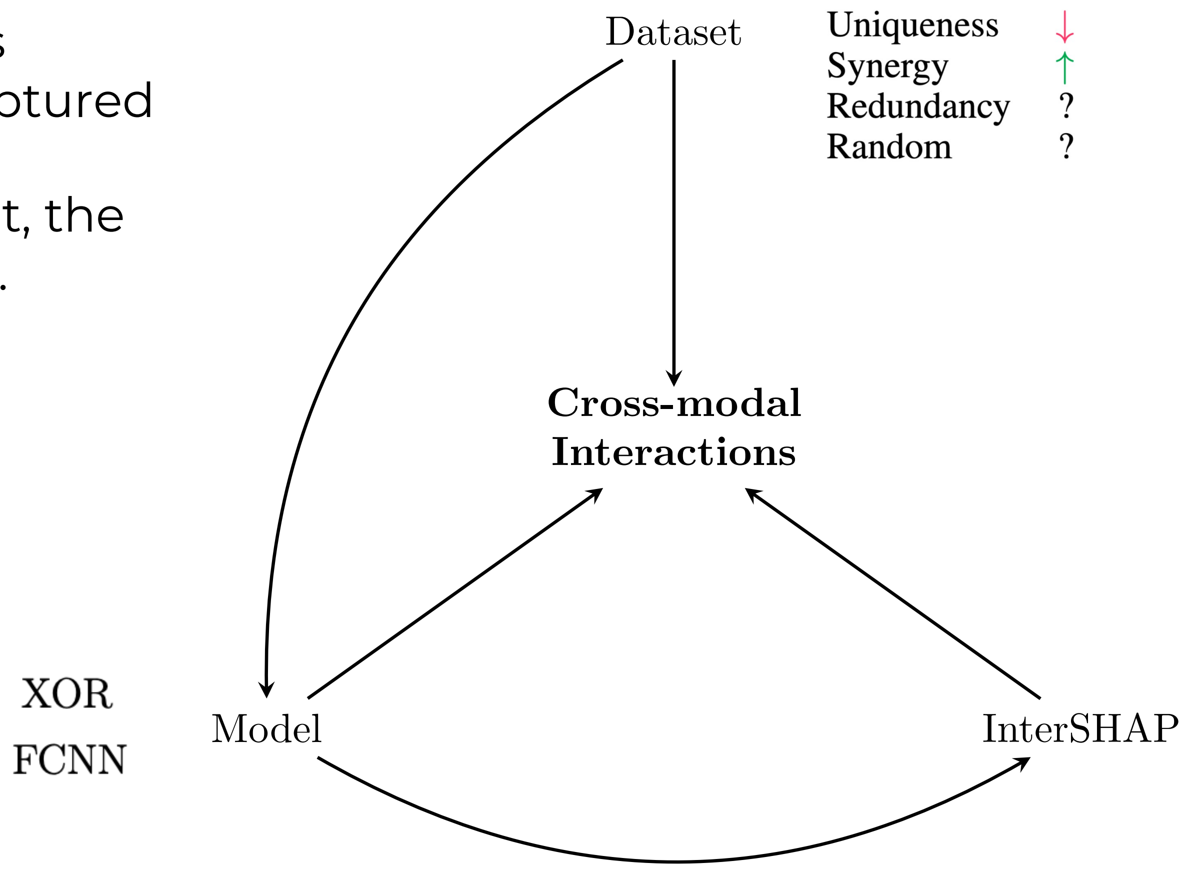
$$\Phi_{ij} = \left| \frac{1}{N} \sum_{a=1}^{N} \phi_{ij}\left(\left(m_1^a, \ldots, m_M^a\right), f\right) \right|, \quad i, j \in \{1, \ldots, M\} \tag{1}$$

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1M} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{M1} & \Phi_{M2} & \cdots & \Phi_{MM} \end{bmatrix} \tag{2}$$

$$InterSHAP = \frac{Interactions}{Behaviour} \tag{3}$$

Verification on Synthetic Data

Three main factors influencing the captured cross-modal interactionsdataset, the model, and metric.

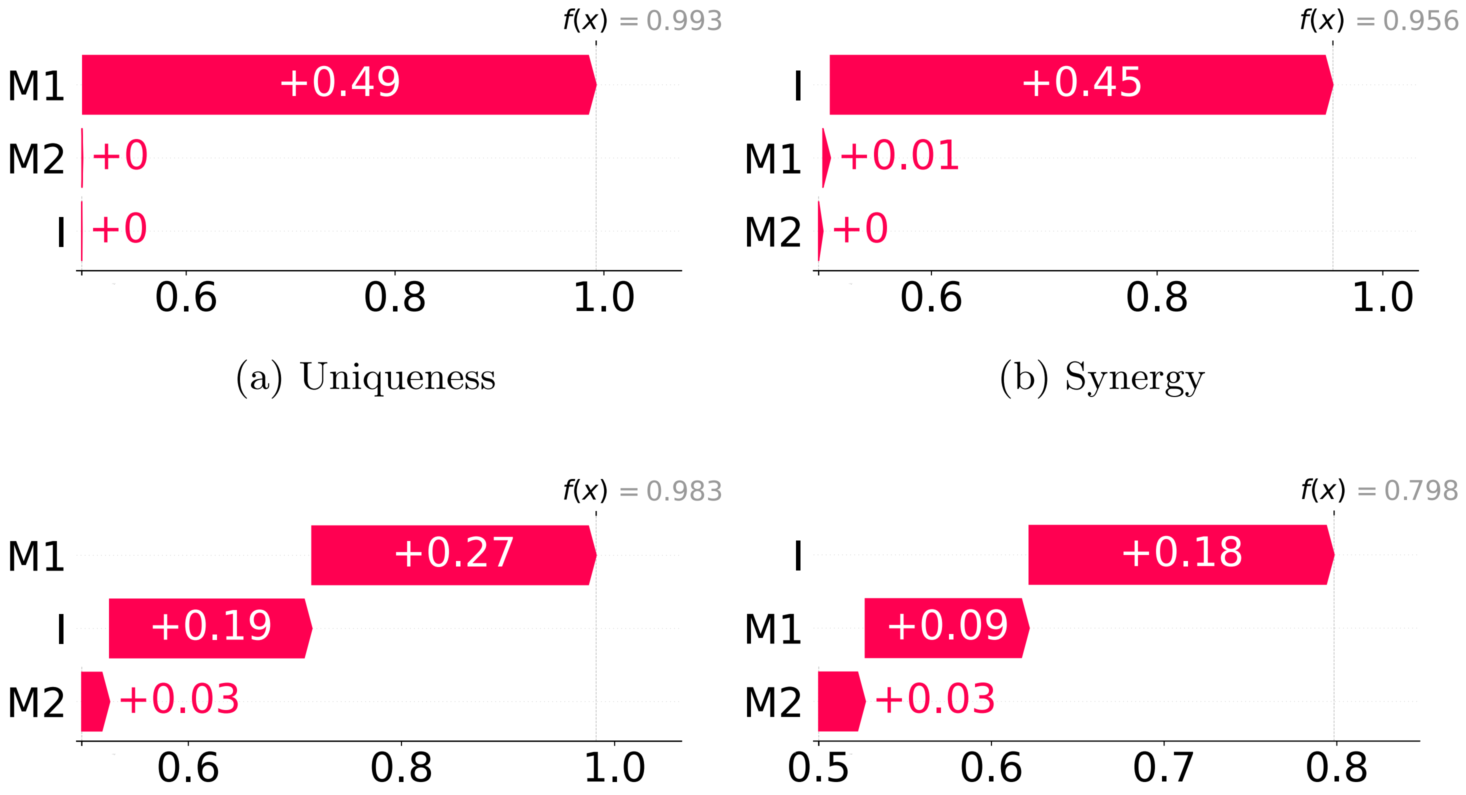


Results – 2 Modalities

InterSHAP values are presented as percentages for both the XOR function and FCNN with early fusion on the HD-XOR datasets. Results for XOR align with expectations, confirming the effectiveness of InterSHAP. For the FCNN, slightly higher values for uniqueness and lower values for synergy suggest the FCNN model did not fully capture all underlying cross-modal interactions from the dataset.

	Uniqueness		Synergy		Redundancy	Random
	XOR	FCNN	XOR	FCNN	FCNN	FCNN
InterSHAP	0.0	0.2 ± 0.1	99.7	98.0 ± 0.5	38.6 ± 0.5	57.8 ± 1.1

Visualisation of Results



(a) Uniqueness (b) Synergy

(c) Redundancy (d) Random

Visualisation of results with FCNN with early fusion on the HD-XOR datasets. M1 represents modality 1, M2 modality 2, and I interactions...

Results > 2 Modalities

	Uniqueness	Synergy	Redundancy
2 Modalities	0.2 ± 0.1	98.0 ± 0.5	38.6 ± 0.5
3 Modalities	0.6 ± 0.2	88.8 ± 0.5	51.9 ± 0.3
4 Modalities	1.2 ± 0.1	64.1 ± 0.8	40.2 ± 0.2

Limitations

- Runtime: $O(N^M)$

Application to healthcare domain

Single Cell Dataset [4]

Modalities
- RNA
- Protein

Task: Cell Class Classification

Table 7: Details of Single Cell.

Class Distribution			
Neutrophil	Erythrocyte	B-Lymphocytes	Monocyte
43.7%	49.8%	1.4%	5.1%

MIMIC III [5]

Modalities
- 12 physiological measurements (e.g. heart rate, 24h)
- static information on patients

Tasks: ICD and Mortality Classification

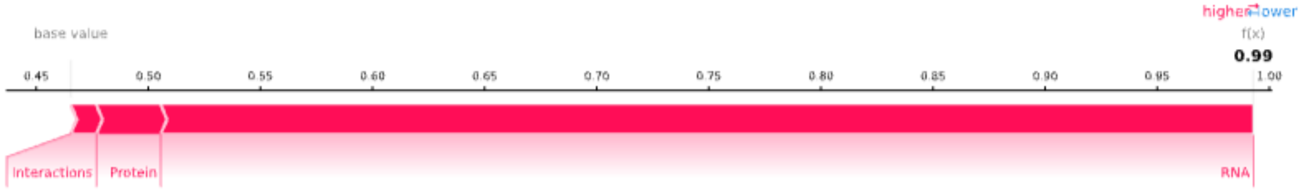
Table 8: Details of MIMIC III, ICD 1 and mortality.

Class Distribution							
ICD		Mortality					
No	Yes	1d	2d	3d	7d	1 year	> 1 year
82.5%	17.5%	76.0%	0.4%	1.3%	1.0%	11.0%	10.3%

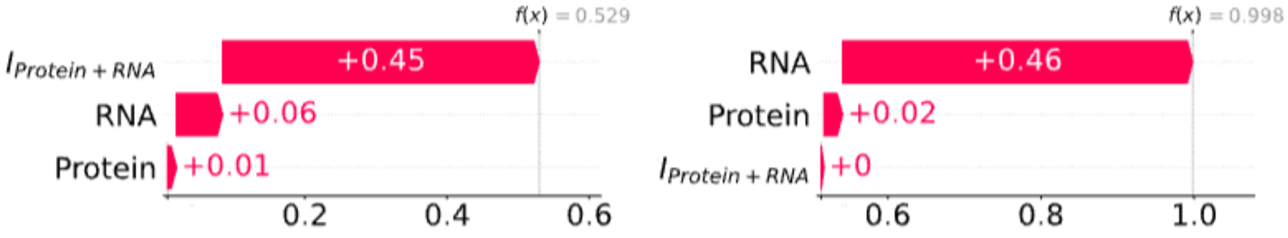
Single Cell Dataset [4]

Table 4: Cross-modal interactions scores on the multimodal single-cell dataset for FCNN with early, intermediate and late fusion. InterSHAP aligns with other SOTA methods, capturing the decline in cross-modal information from early to late fusion.

	Single-Cell		
	early	intermediate	late
InterSHAP	1.9 ± 0.4	1.5 ± 0.4	0.4 ± 0.1
PID	0.08 ± 0.01	0.08 ± 0.01	0.06 ± 0.0
EMAP_{gab}	0 ± 0	0 ± 0	0 ± 0
SHAPE	1.0 ± 0.2	0.7 ± 0.2	0 ± 0

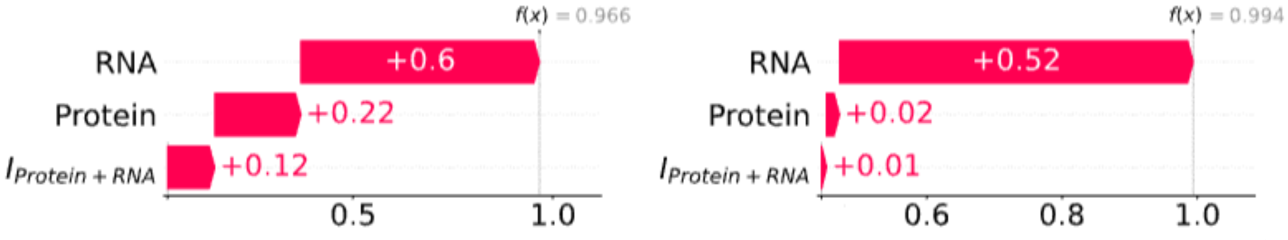


(a) Explanation over the whole dataset



(b) B-Cell Progenitor

(c) Erythrocyte Progenitor



(d) Monocyte Progenitor

(e) Neutrophil Progenito

MIMIC III [5]

Table 5: Cross-modal interactions scores on the MIMIC III dataset for baseline model and MVAE model from Multi-Bench implementation. InterSHAP aligns with other SOTA methods, capturing greater cross-modal interaction in the baseline compared to MVAE, while uniquely quantifying the proportional contribution of cross-modal interactions.

	ICD-9		Mortality	
	baseline	MVAE	baseline	MVAE
InterSHAP	1.2 ± 0.2	6.8 ± 1.3	11.0 ± 0.5	12.3 ± 2.8
PID	0.06 ± 0.01	0.09 ± 0.01	0.10 ± 0.01	0.11 ± 0.01
EMAP_{gap}	0 ± 0	1.2 ± 0.0	-0.8 ± 0.1	0.9 ± 0.1
SHAPE	0.2 ± 0	0.6 ± 0	0.2 ± 0.2	0.7 ± 0.2

Summary

- Novel cross-modal interaction score: **InterSHAP -** Open-Source implementation with integration into SHAP package
  - >2 Modalities
  - Local
  - Unsupervised
  - Performance agnostic
- Application to healthcare multimodal datasets

More in our paper:

- Quantitative evaluation of existing cross-modal interaction scores

References

[1] Hessel, J.; and Lee, L. 2020. Does my multimodal model learn cross-modal interactions? It's harder to tell than you might think! In Proceedings of the 2020 Conference on Em- pirical Methods in Natural Language Processing (EMNLP), 861–877. Association for Computational Linguistics.

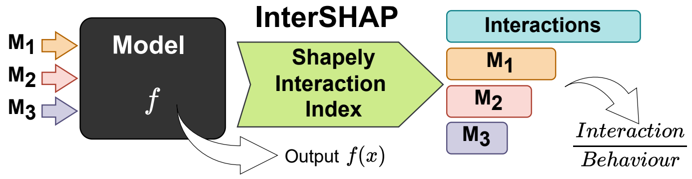
[2] Hu, P.; Li, X.; and Zhou, Y. 2022. SHAPE: An Unified Approach to Evaluate the Contribution and Cooperation of Individual Modalities. In Proceedings of the Thirty-First Inter- national Joint Conference on Artificial Intelligence, 3064– 3070. Vienna, Austria: International Joint Conferences on Artificial Intelligence Organization.

[3] Liang, P. P.; Cheng, Y.; Fan, X.; Ling, C. K.; Nie, S.; Chen, R.; Deng, Z.; Allen, N.; Auerbach, R.; Mahmood, F.; Salakhutdinov, R. R.; and Morency, L.-P. 2023. Quantifying & Modeling Multimodal Interactions: An Information Decomposition Framework. Advances in Neural Information Processing Systems, 36: 27351–27393.

[4] Burkhardt, D.; Luecken, M.; Benz, A.; Holderrieth, P.; Bloom, J.; Lance, C.; Chow, A.; and Holbrook, R. 2022. Open Problems - Multimodal Single-Cell Integration.

[5] Liang, P. P.; Lyu, Y.; Fan, X.; Wu, Z.; Cheng, Y.; Wu, J.; Chen, L.; Wu, P.; Lee, M. A.; Zhu, Y.; Salakhutdinov, R.; and Morency, L.-P. 2021. MultiBench: Multiscale Benchmarks for Multimodal Representation Learning. In Vanschoren, J.; and Yeung, S.-K., eds., Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks 1, NeurIPS Datasets and Benchmarks 2021, December 2021.

Shapley Interaction Index



$$\phi_{ij}(M,f) = \sum_{S \subseteq M \setminus \{i,j\}} \frac{|S|!(M-|S|-2)!}{2(M-1)!} \nabla ij(S,f), \quad i \neq j.$$

$$\nabla ij(S,f) \; = \; \left[f_{S\cup\{ij\}}\left(S \cup \{ij\}\right) - f_{S\cup\{i\}}\left(S \cup \{i\}\right) - f_{S\cup\{j\}}\left(S \cup \{j\}\right) + f_S(S) \right]$$

$$\phi_{ii}(M,f) = \phi_i(M,f) - \sum_{j \in M} \phi_{ij}(M,f) \quad \forall i \neq j.$$

$$\phi_i(M,f) = \sum_{S \subseteq M \setminus \{i\}} \frac{|S|!(|M|-|S|-1)!}{|M|!} \Delta$$

$$\Delta = \left[f_{S\cup\{i\}}\left(S \cup \{i\}\right) - f_S\left(S\right) \right].$$