MultiModN- Multimodal, Multi-Task, Interpretable Modular Networks

VINITRA SWAMY*, MALIKA SATAYEVA*, JIBRIL FREJ, THIERRY BOSSY, THIJS VOGELS, MARTIN JAGGI, TANJA KÄSER*, MARY-ANNE HARTLEY*

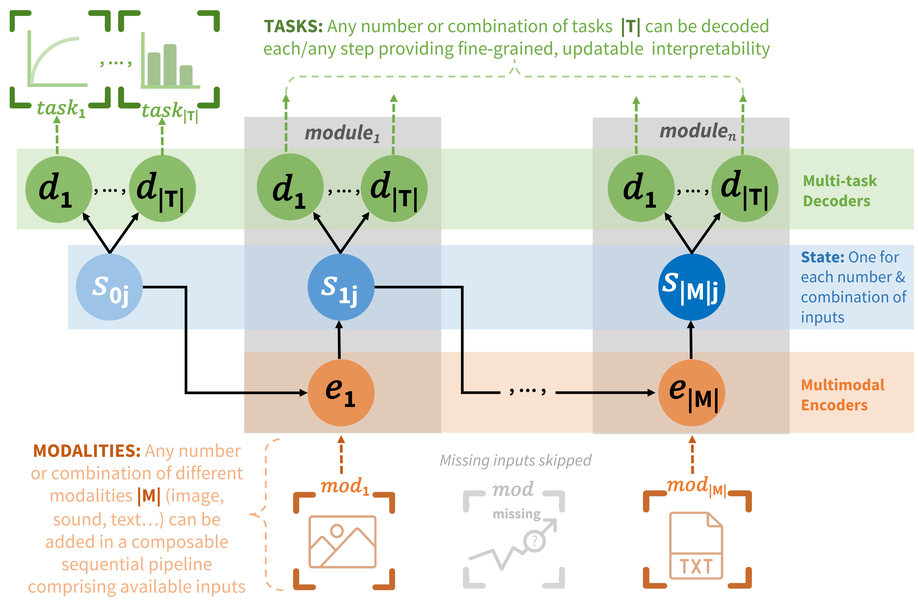
MultiModN is a multimodal, modular network that fuses latent representations in a sequence of any number, combination, or type of modality while providing granular real-time predictive feedback on any number or combination of predictive tasks

NEURAL INFORMATION PROCESSING SYSTEMS

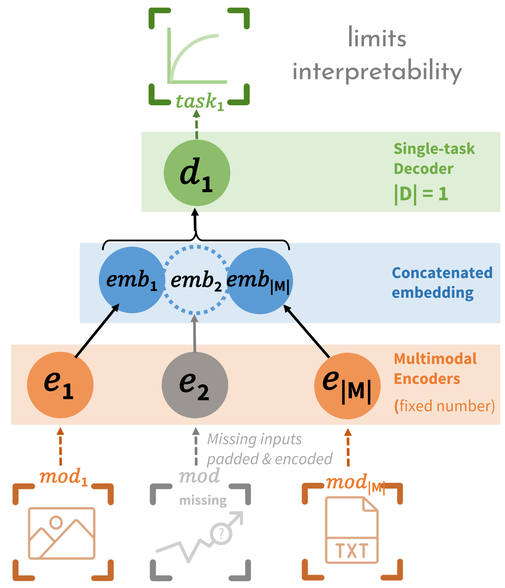
EPFL Yale

github.com/epfl-iglobalhealth/MultiModN

MultiModN



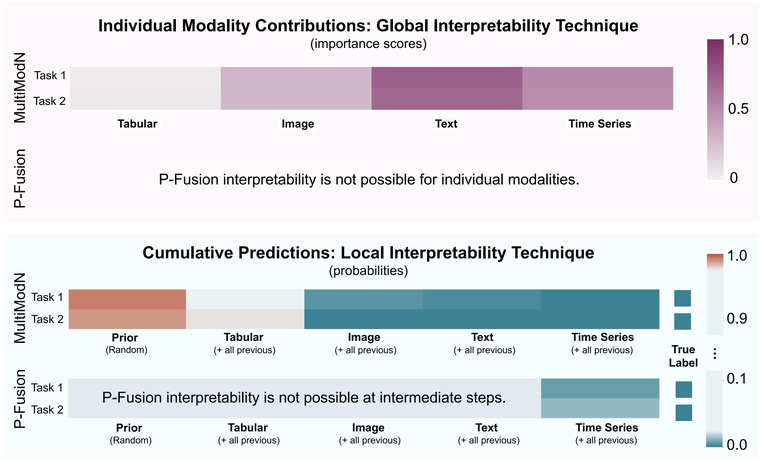
Monolithic P-Fusion



dependent on modality availability

Interpretable-by-design

MultiModN has inherent modality-specific global (IMC) and local (CP) model explainability.



Using IMC, we identify that the text modality has the most importance in the MIMIC dataset. For CP, we can measure which modality input causes the decision to flip from a negative (Prior) to a neutral (Tabular) to a positive (Image) diagnosis.

MultiModN

1) matches parallel MM fusion (P-Fusion) for a range of real-world tasks

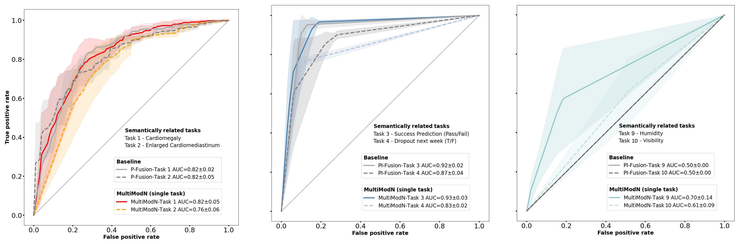
2) composable at inference

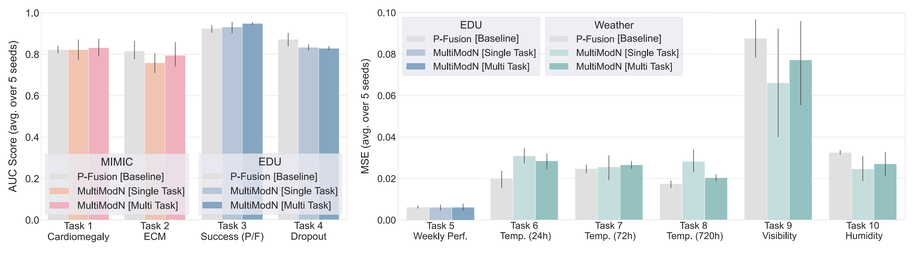
3) robust to the bias of missing not-at-random (MNAR) modalities

4) inherently interpretable

5) easily extended to any number or combination of tasks

Evaluated on 10 real world tasks across multimodal healthcare, education, and weather benchmarks

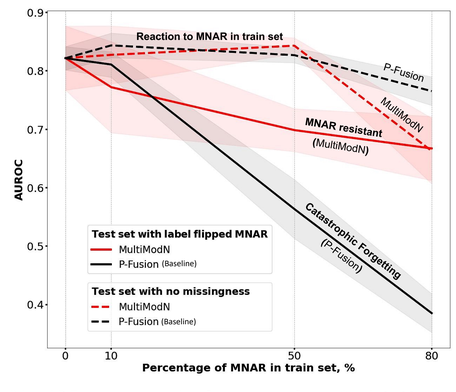




MultiModN has the advantage of being naturally extensible to the prediction of multiple tasks without negatively impacting the performance of individual tasks.

Robust to Missingness

MultiModN is robust to bias from missing input modalities (catastrophic MNAR failure).



MultiModN and P-Fusion are trained on four versions of MIMIC with 0–80% MNAR. They are tested on a test set with no MNAR missingness (- - -) or a test set where the biased missingness is label-flipped i.e. MNAR occurs in the other binary class as compared with the train (—). MultiModN is MNAR resistant, while P-Fusion exhibits catastrophic forgetting.