Deep Learning for Medical Image Analysis

COMP5423

Hao CHEN

Dept. of CSE,CBE&LIFS, HKUST

jhc@cse.ust.hk

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

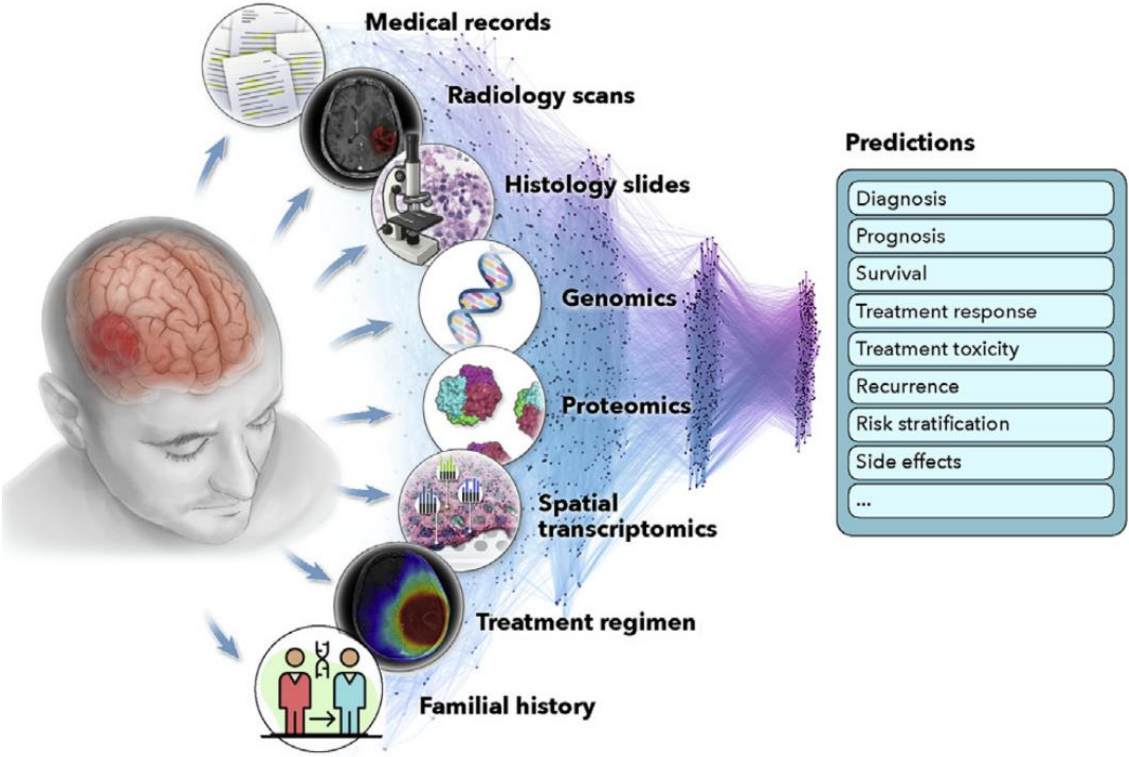
香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Multimodal Learning in Healthcare

- Introduction
- Methods
- Interpretability
- Multimodal data interconnection
- Challenge and clinical adoption

Multimodal Data Integration

- The **aim** is to extract and combine complementary contextual information across different modalities for better decision-making.
- Ranging from **radiology, histology, clinical and laboratory tests, to familial and patient histories**.

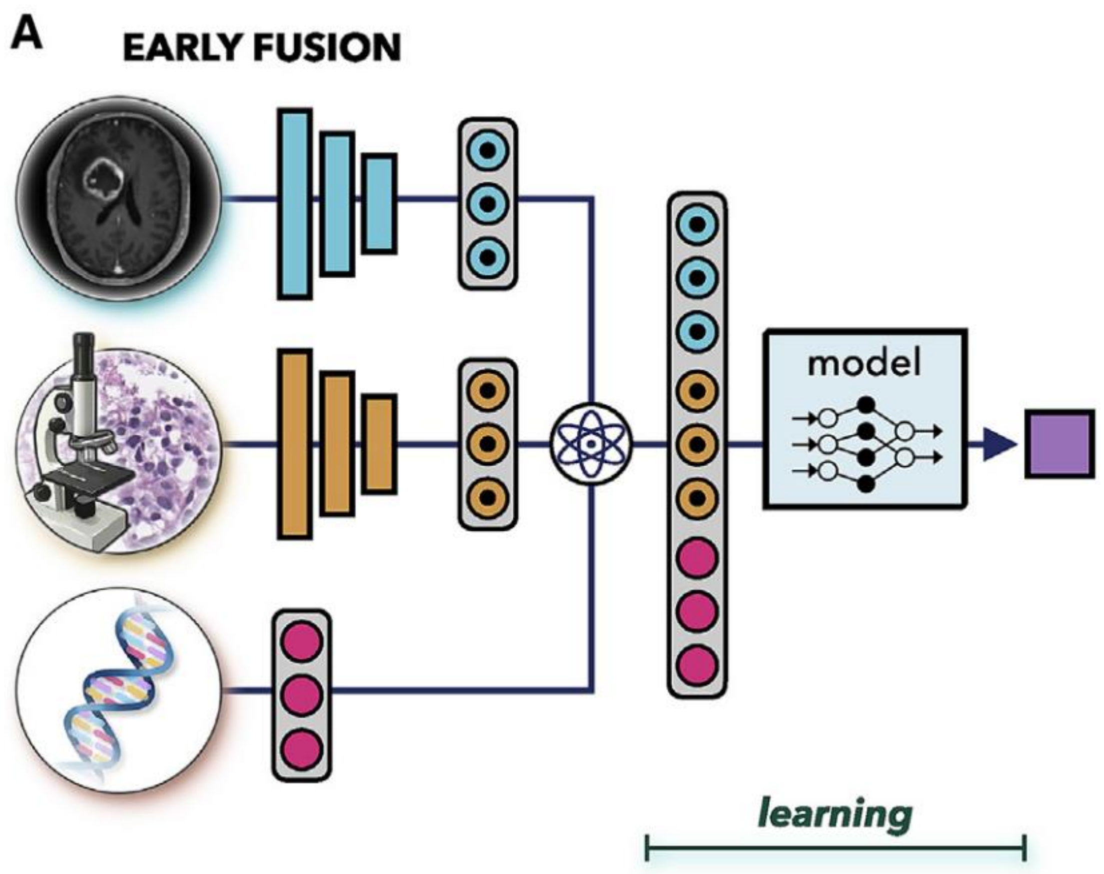


J. Lipkova et al., “Artificial intelligence for multimodal data integration in oncology,” Cancer Cell, vol. 40, no. 10, pp. 1095–1110, Oct. 2022, doi: https://doi.org/10.1016/j.ccell.2022.09.012.

Multimodal Learning in Healthcare

- Introduction
- Methods
- Interpretability
- Multimodal data interconnection
- Challenge and clinical adoption

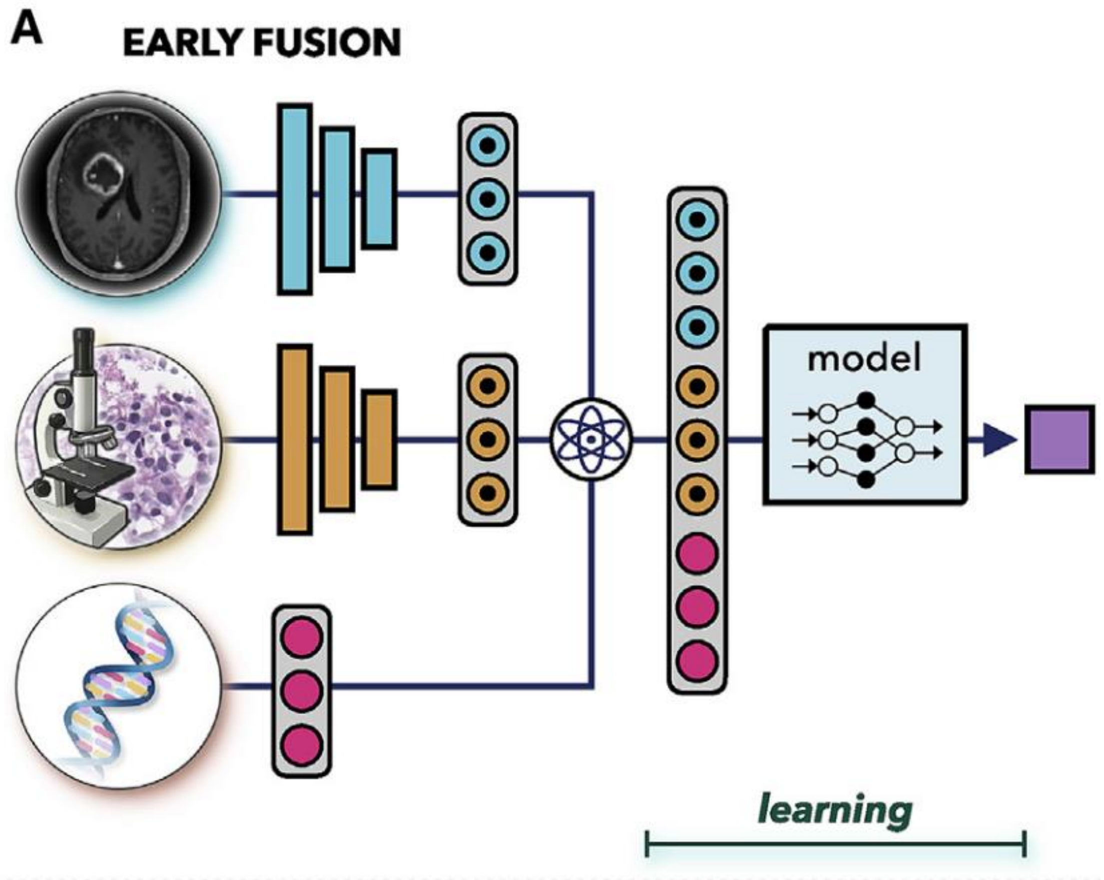
Early Fusion



- Integrate information from all modalities **at the input level** before feeding it into a single model.
- The joint representation is built through operations such as **vector concatenation**, **element-wise sum, element-wise multiplication**, or **bilinear pooling**.

Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., and Lungren, M.P. (2020). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit. Med. 3, 136–139.
Ramachandram, D., and Taylor, G.W. (2017). Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process. Mag. 34, 96–108.

Early Fusion



- **Only one model** is trained.
- Assumed that the single model is **well suited** to all modalities.
- Require a certain level of **alignment or synchronization**.

Scope of application

- If the modalities come from significantly different time points, then early fusion might not be an appropriate choice.

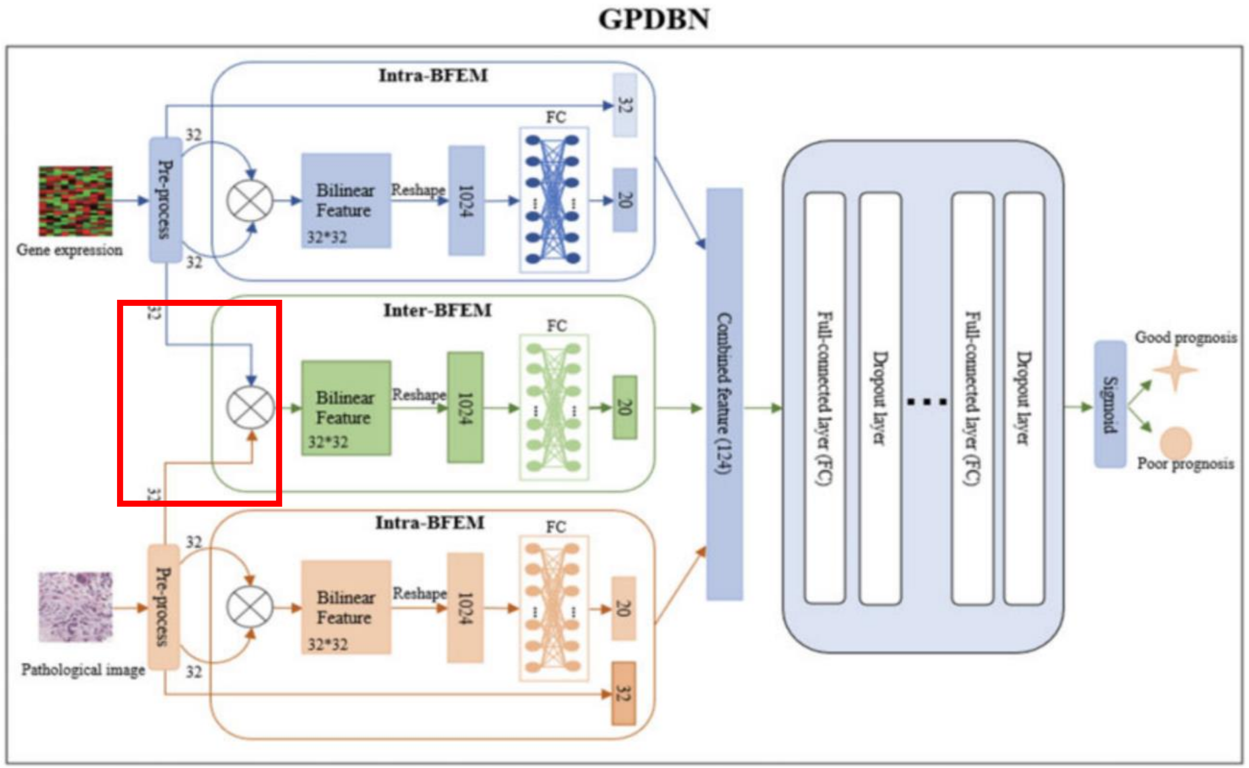
Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., and Lungren, M.P. (2020). Fusion of medical imaging and elec- tronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit. Med. 3, 136–139.
Ramachandram, D., and Taylor, G.W. (2017). Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process. Mag. 34, 96–108.

Early Fusion

GPDBN

Inter-BFEM utilizes a bilinear function of g and p from genomic data and pathological images.

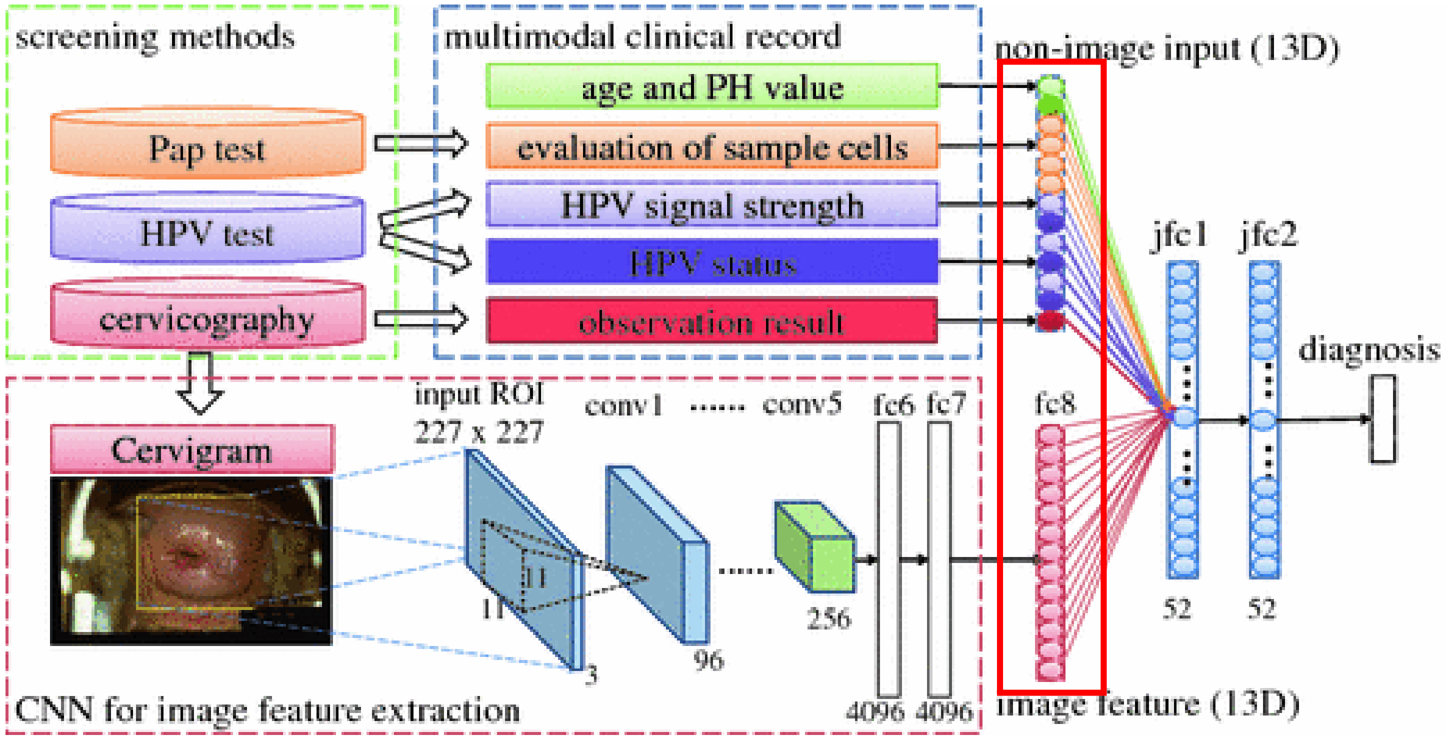
$$f_k^{inter} = ReLU(W_k vec(\boldsymbol{g}\boldsymbol{p}^T) + b_k)$$



Wang Z, Li R, Wang M, et al. GPDBN: deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction[J]. Bioinformatics, 2021, 37(18): 2963-2970.

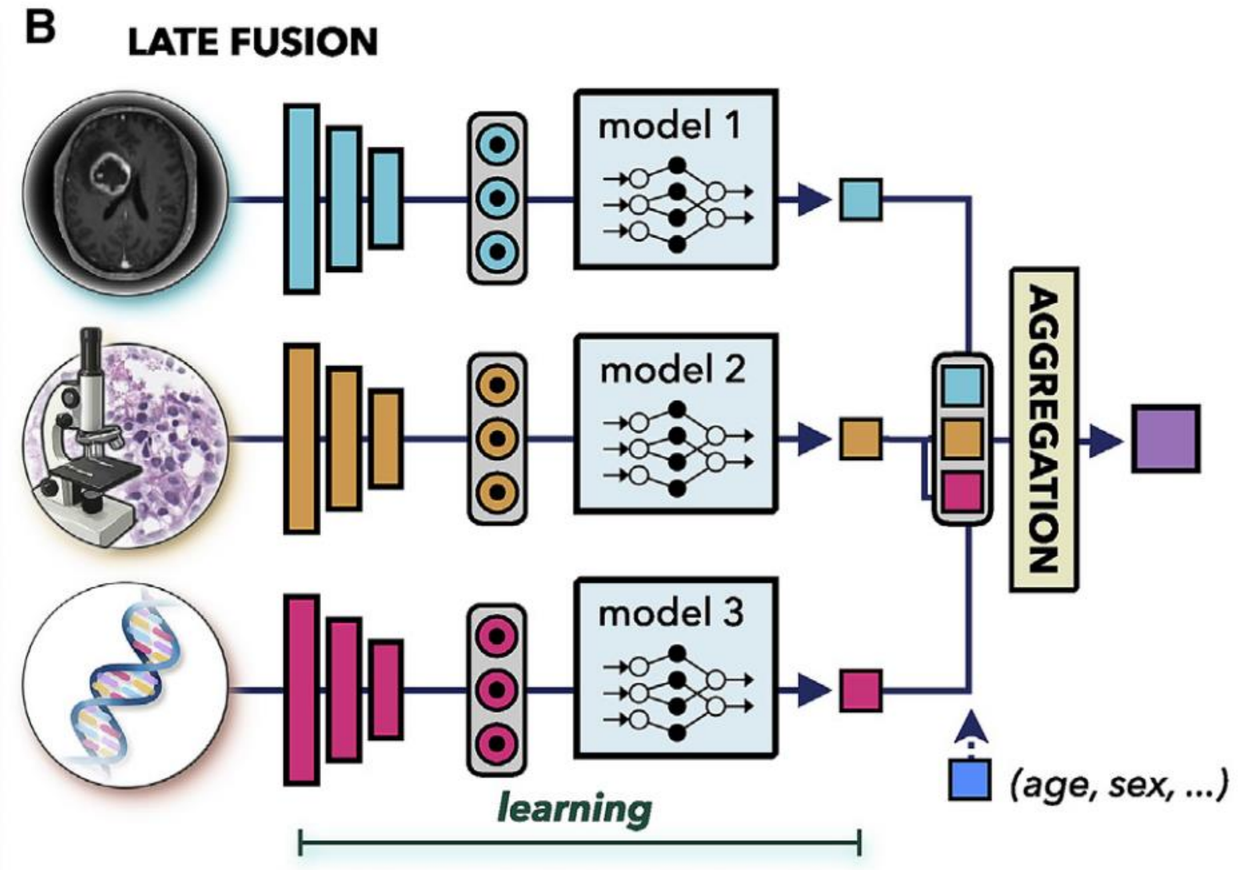
Early Fusion

Fusion of a cervigram and EMRs for cervical dysplasia diagnosis



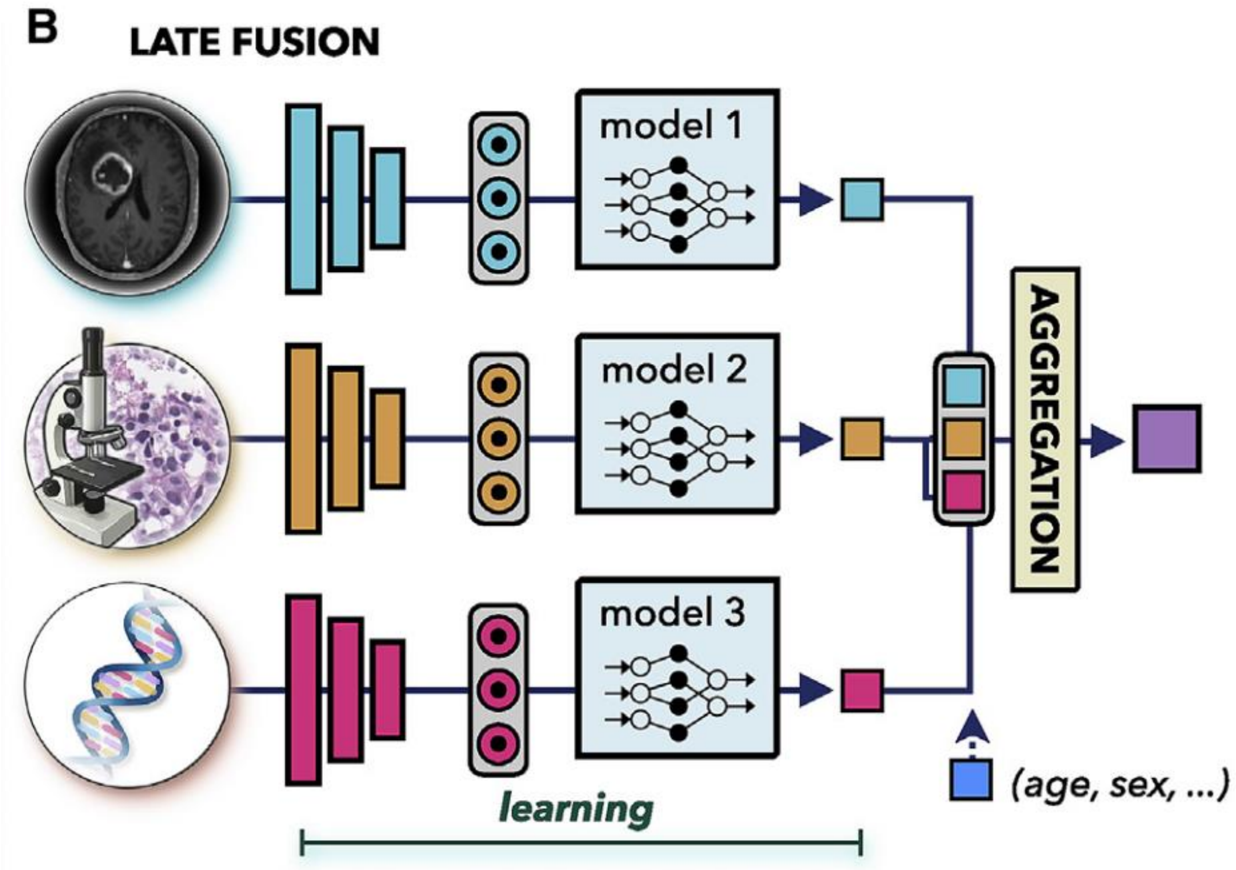
Xu T, Zhang H, Huang X, et al. Multimodal deep learning for cervical dysplasia diagnosis[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19. Springer International Publishing, 2016: 115-123.

Late fusion



- Also known as **decision-level fusion**
- Train a **separated** model for each modality and aggregate the **predictions from individual models** for the final prediction.

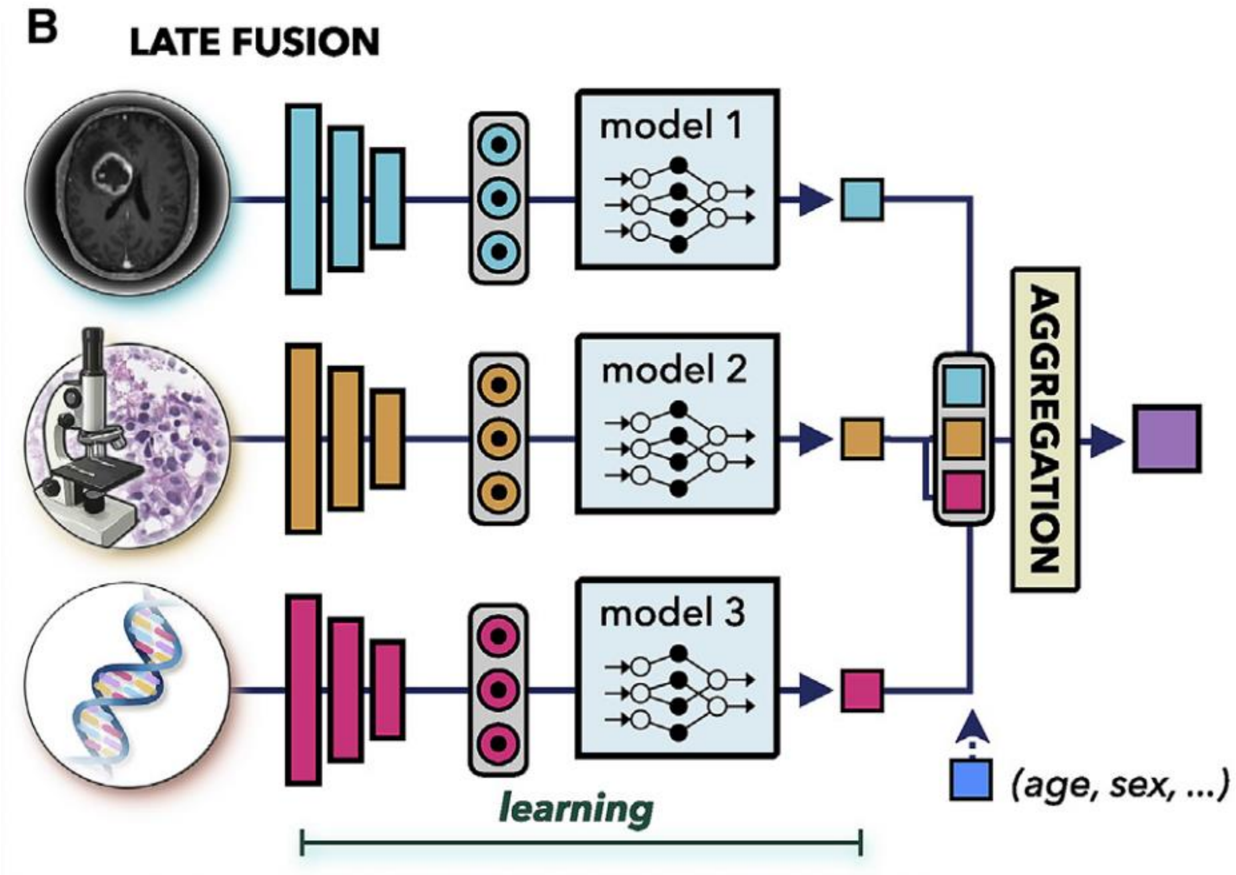
Late fusion



Pros.

- Allow one to use a **different model architecture** for each modality.
- Do not pose any constraints on data **synchronization**.
- In cases of **missing or incomplete data**, late fusion retains the ability to make predictions.
- Individual models tend to be uncorrelated, resulting in potentially **lower bias** and variance in late-fusion predictions.

Late fusion



Cons.

- In situations when **information density varies significantly** across modalities, predictions can be **heavily influenced by the most dominant modality**.

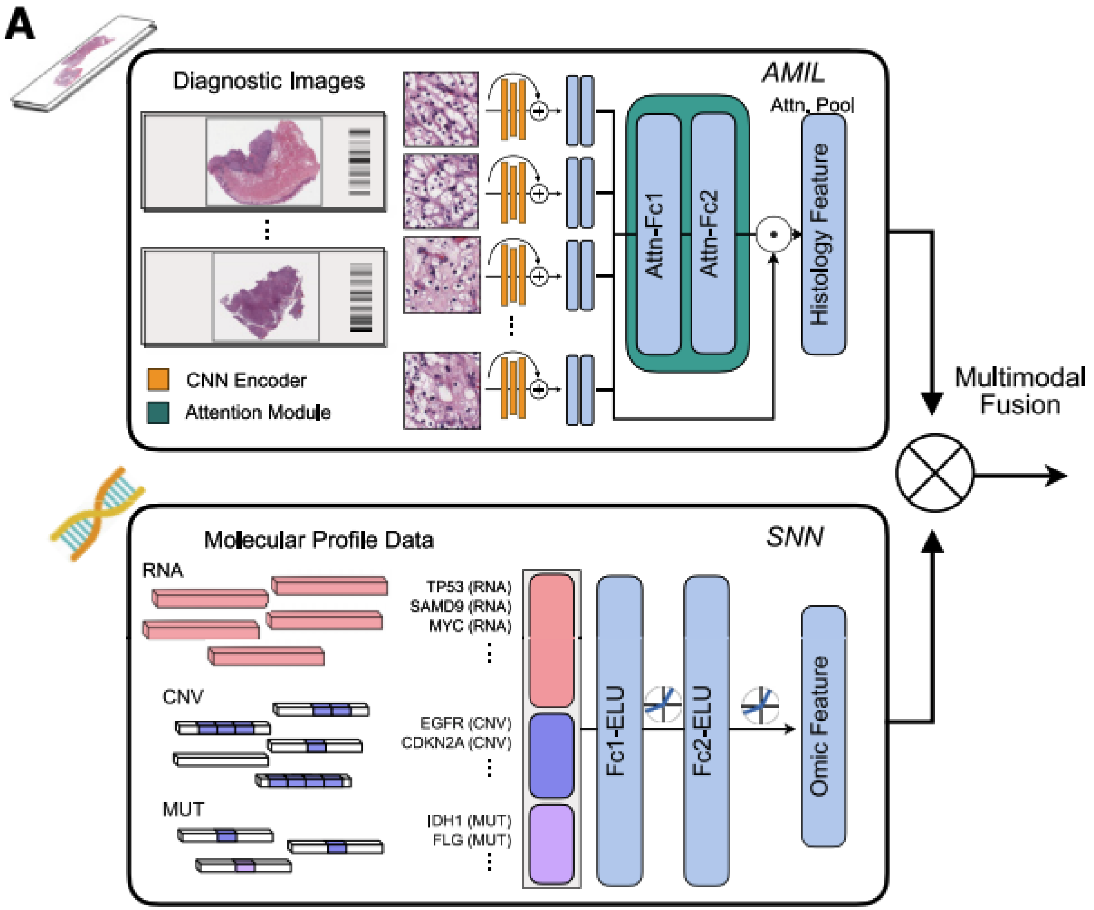
Late fusion

Scope of applications

- Suitable for systems with **large data heterogeneity** or modalities from different time points.
- **Inclusion of a new modality** can be performed without the need to retrain the full model.
- **Simple covariates**, such as age or gender, are often included through late fusion due to its simplicity.
- If the unimodal data **does not have strong inter-dependencies**, late fusion might be preferable thanks to the simpler architecture and smaller number of parameters.

Late fusion

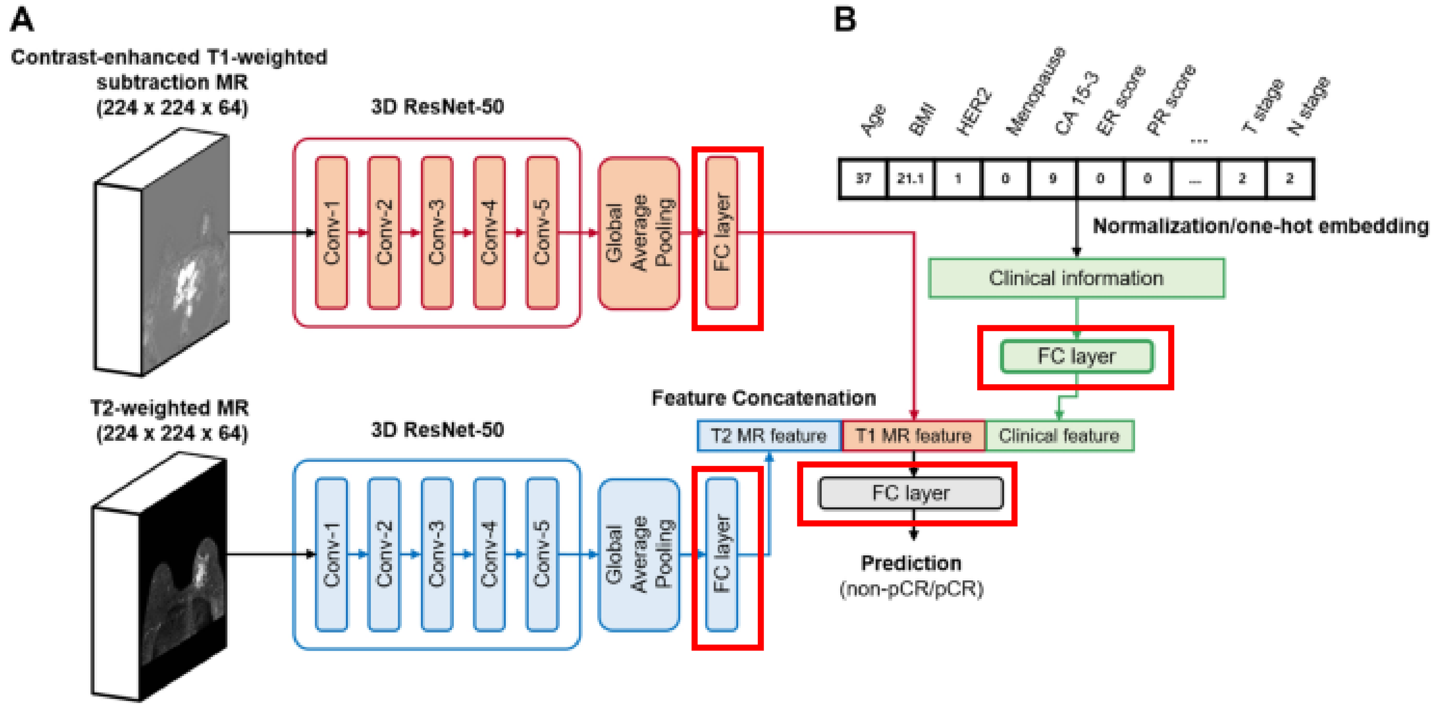
- An attention-based multiple instance learning (**AMIL**) network for processing **WSIs**.
- A self-normalizing network (**SNN**) for processing **molecular data features**.
- Multimodal **fusion** layer that computes the **Kronecker Product** to model pairwise feature interactions.



R. J. Chen et al., "Pan-cancer integrative histology-genomic analysis via multimodal deep learning," Cancer Cell, vol. 40, no. 8, pp. 865-878.e6, Aug. 2022, doi: 10.1016/j.ccell.2022.07.004.

Late fusion

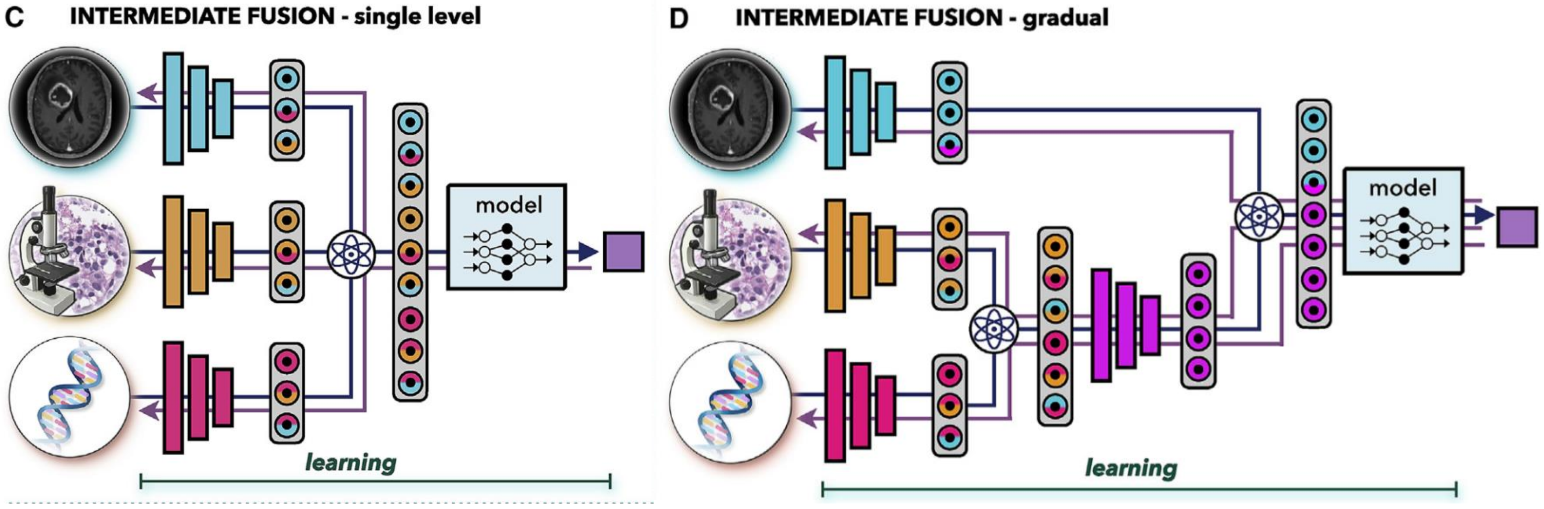
- Two 3D ResNet-50 for contrast-enhanced T1W subtraction MR images and T2W MR images.
- FC layer was used for clinical information inputs.

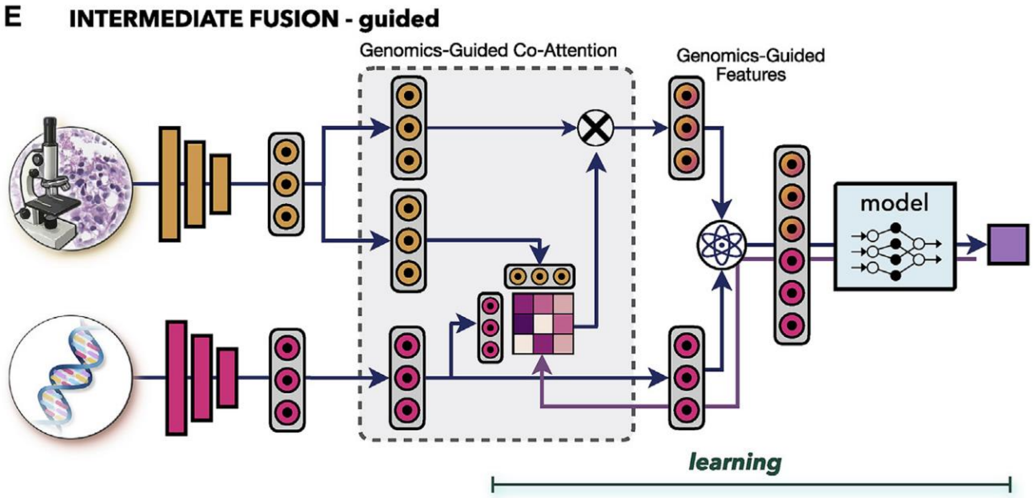


Joo, S., Ko, E.S., Kwon, S. et al. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. Sci Rep 11, 18800 (2021).

Intermediate fusion

- Single level fusion
- Gradual fusion
- Guided fusion

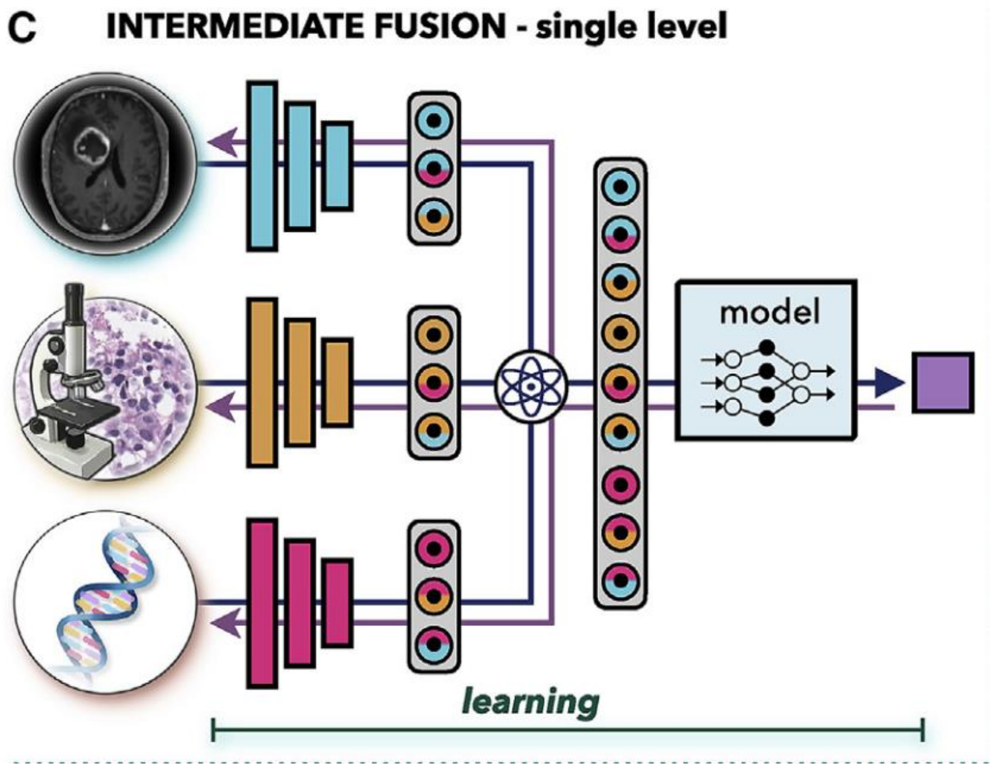




- Can combine individual modalities at **different levels of abstractions**.

Intermediate fusion

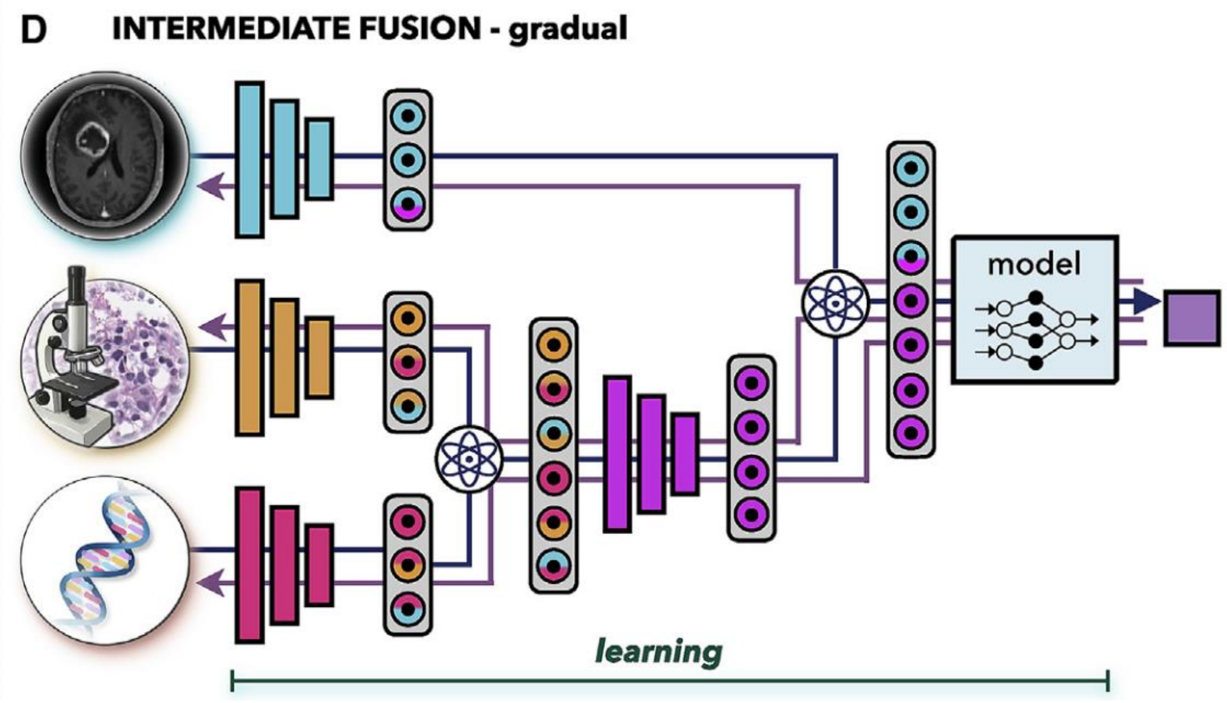
Single level fusion



- **Different from early fusion**:
  in early fusion the unimodal embeddings are not affected by the multimodal context (**gradient backward**).

Intermediate fusion

Gradual fusion

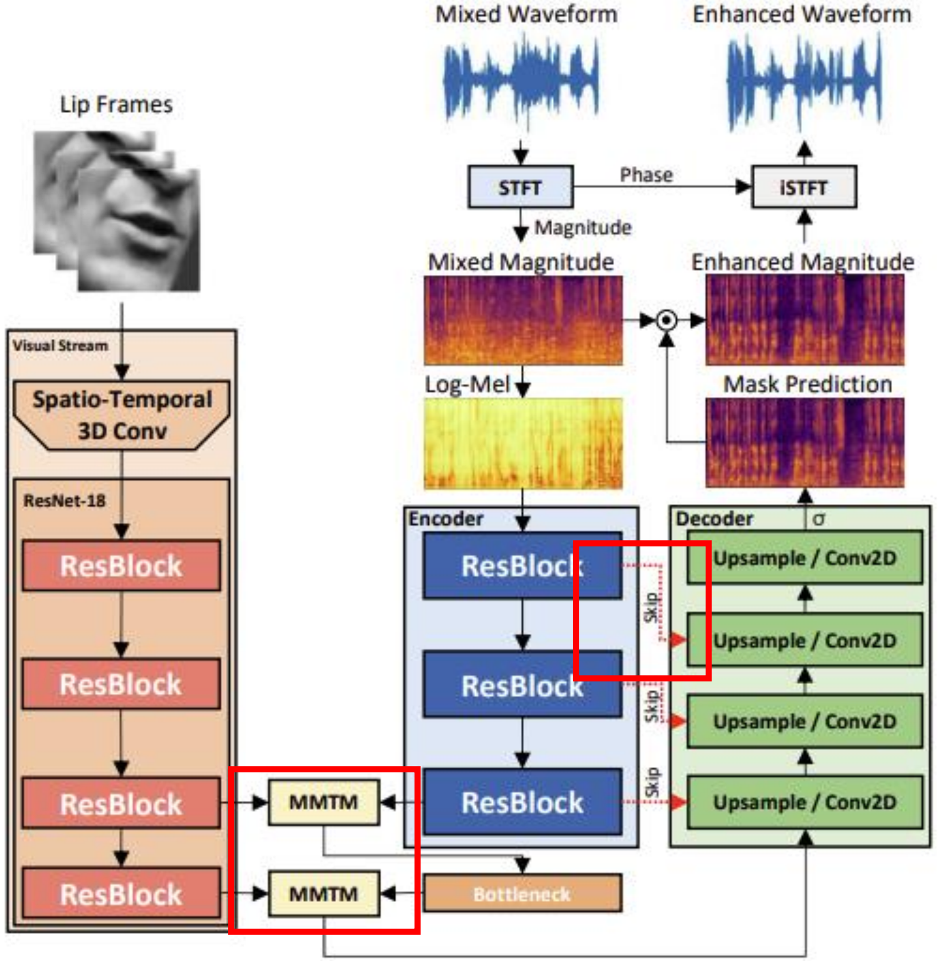


- Allow one to combine data **from highly correlated** channels at the same level, followed by fusion with **less correlated** data in later layers.
- Force the model to consider the **cross-correlations** between specific modalities.

Intermediate fusion

Gradual fusion

Audio-Visual Speech Enhancement (AVSE)

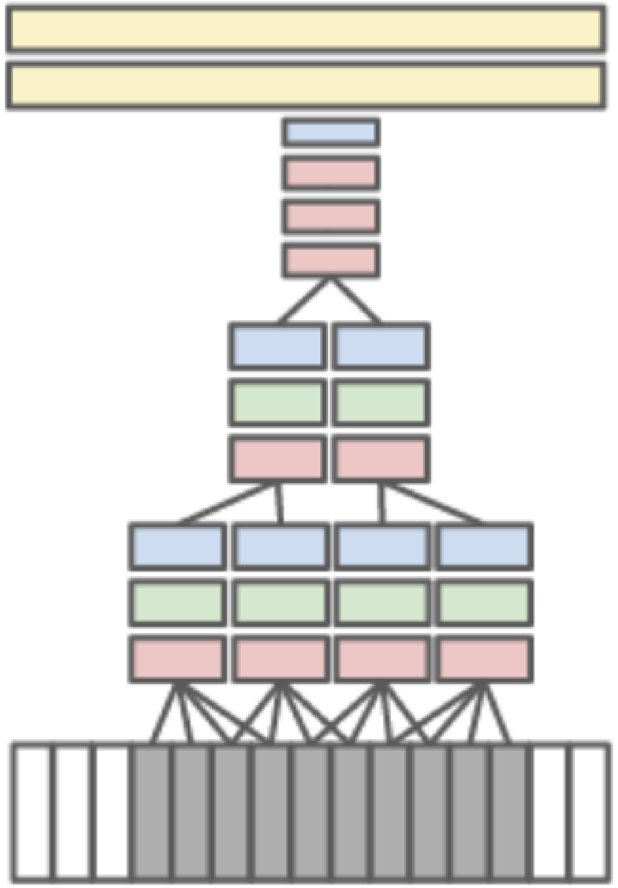


Joze, H.R.V., Shaban, A., Iuzzolino, M.L., and Koishida, K. (2020). Mmtm: multimodal transfer module for cnn fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13289–13299.

Intermediate fusion

Gradual fusion

- Slowly fuse temporal information throughout the network.
- **Higher layers** get access to **progressively more global information** in both spatial and temporal dimensions.

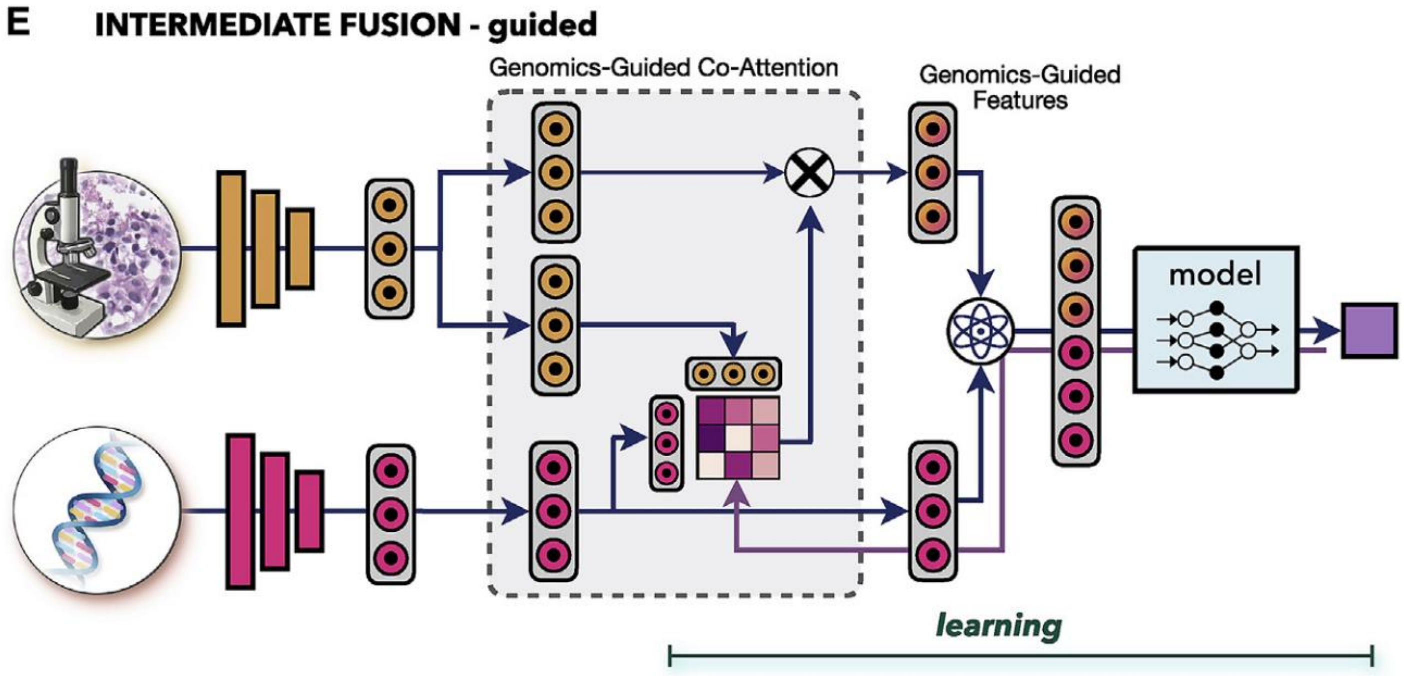


Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1725-1732

Intermediate fusion

Guided fusion

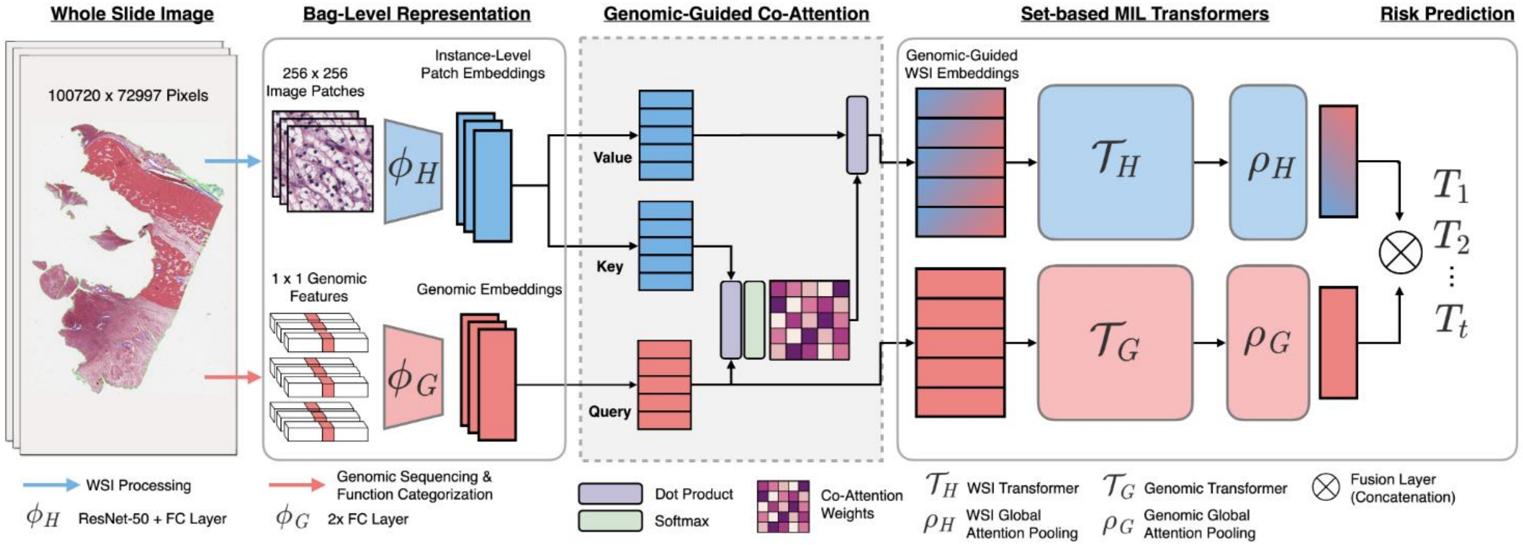
- To use information from one modality to guide feature extraction from another modality.



Intermediate fusion

Guided fusion

- Different **tissue regions** might be relevant in the presence of **specific mutations**.
- Learn **co-attention scores** that reflect the relevance of different histology features in the presence of specific molecular information.



$$\mathrm{CoAttn}_{G\to H}(G,H) = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)$$

R. J. Chen et al., "Multimodal Co-Attention Transformer for Survival Prediction in Gigapixel Whole Slide Images," in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, Oct. 2021, pp. 3995–4005.

Methods

Summary

- There is no conclusive evidence that one fusion type is ultimately better than the others
- Each type is heavily data and task specific.

Early fusion

Late fusion

Intermediate fusion

Multimodal Learning in Healthcare

- Introduction
- Methods
- Interpretability
- Multimodal data interconnection
- Challenge and clinical adoption

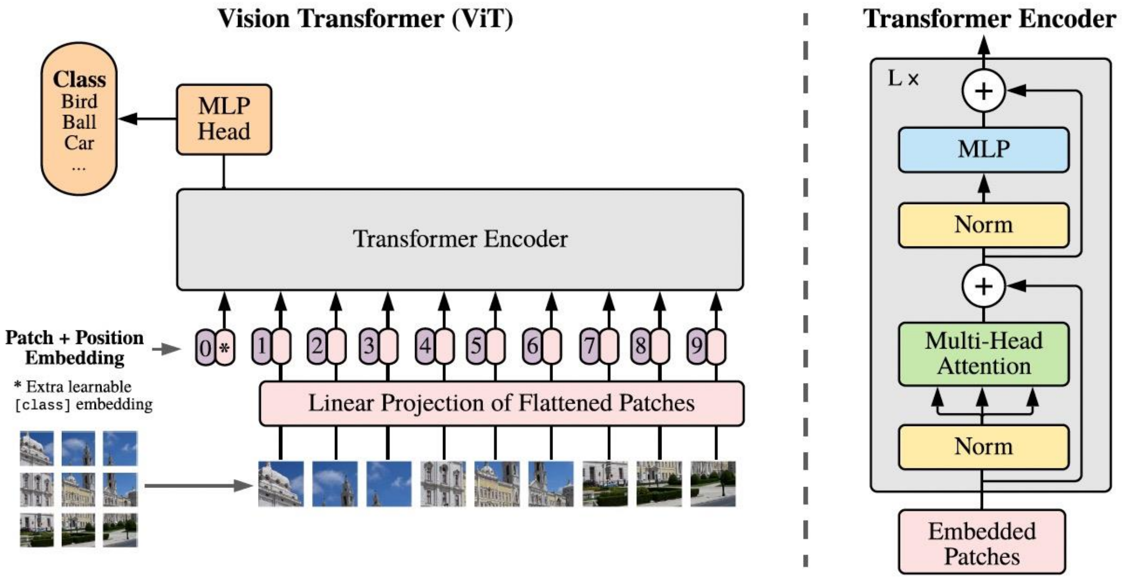
Interpretability

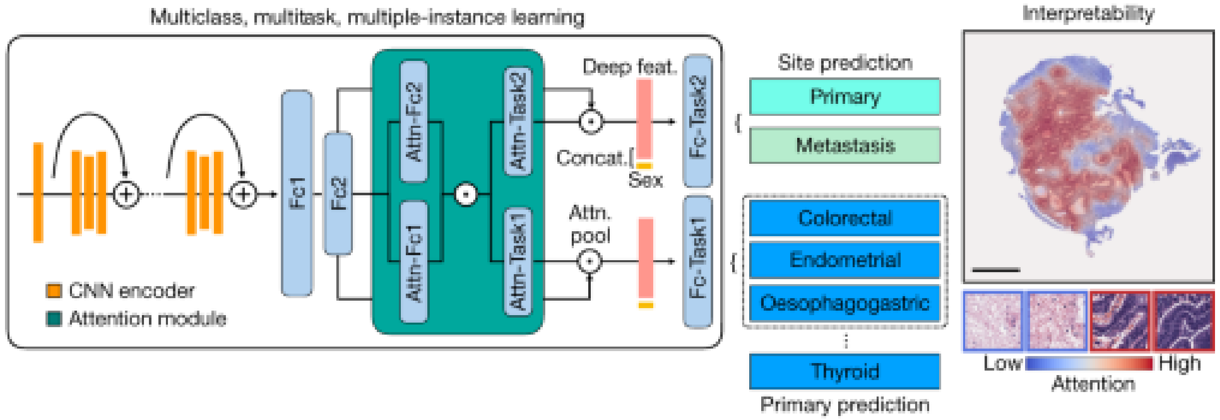
- Interpretability and model introspection is a crucial component of AI development, deployment, and validation.
- AI model is able to learn **abstract feature representations**.
- Sometimes, models might **fail to generalize** when presented with new data, as the models might use **spurious shortcuts** for predictions, instead of learning clinically relevant aspects.
- Interpretable methods allow us to introspect parts of the data **deemed important by the model in making predictive determinations**.
- On the other hand, the models can **discover novel and clinically relevant insights**.

Interpretability

Histopathology

VITs or **MIL** can reveal the relative **importance** of each image **patch** for the model predictions.





$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

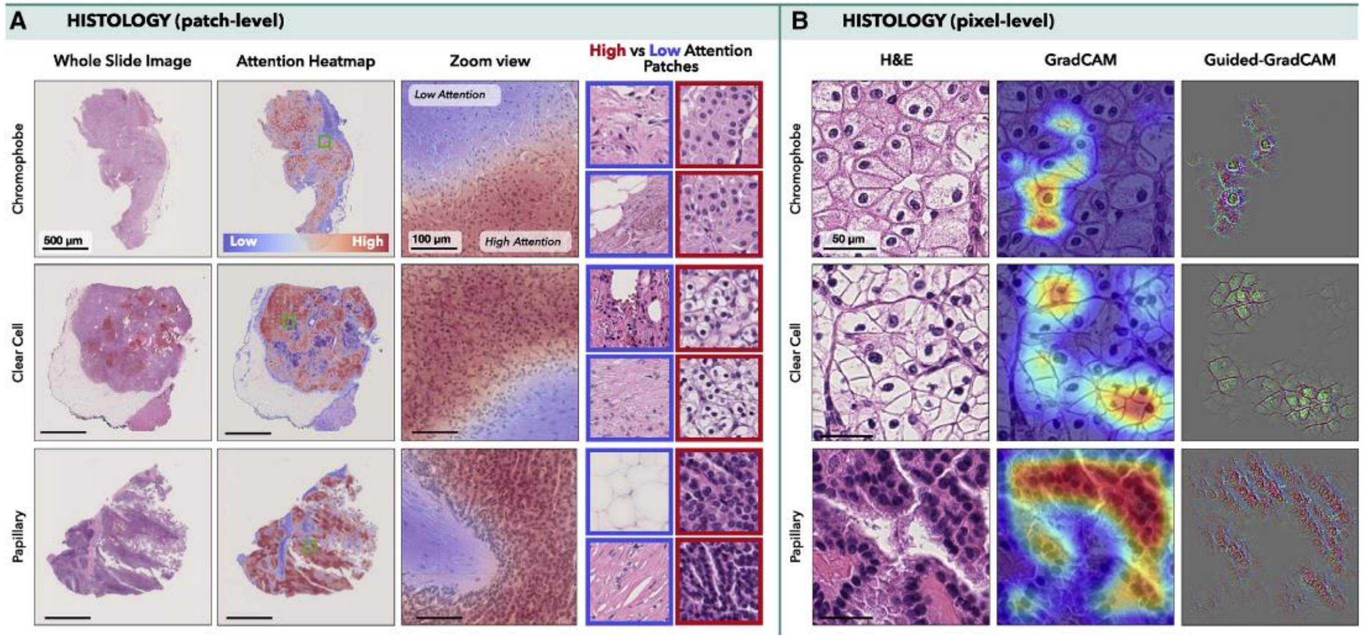
Dosovitskiy, et al. An image is worth 16x16 words: Transformers for image recognition at scale. ICLR 2021.
Lu, M.Y., Chen, T.Y., Williamson, D.F.K., Zhao, M., Shady, M., Lipkova, J., and Mahmood, F. (2021). Ai-based pathology predicts origins for cancers of unknown primary. Nature 594, 106–110.

Interpretability

Histopathology

Depending on the model architecture **attention or probability scores** can be mapped to obtain **slide-level** attention heatmaps.

1. **Attention**
2. **Class activation methods (CAMs):**
   - GradCAM or GradCAM++, guided-GradCAM
   - Allow one to determine the importance of the model inputs (e.g., pixels) for each prediction class.



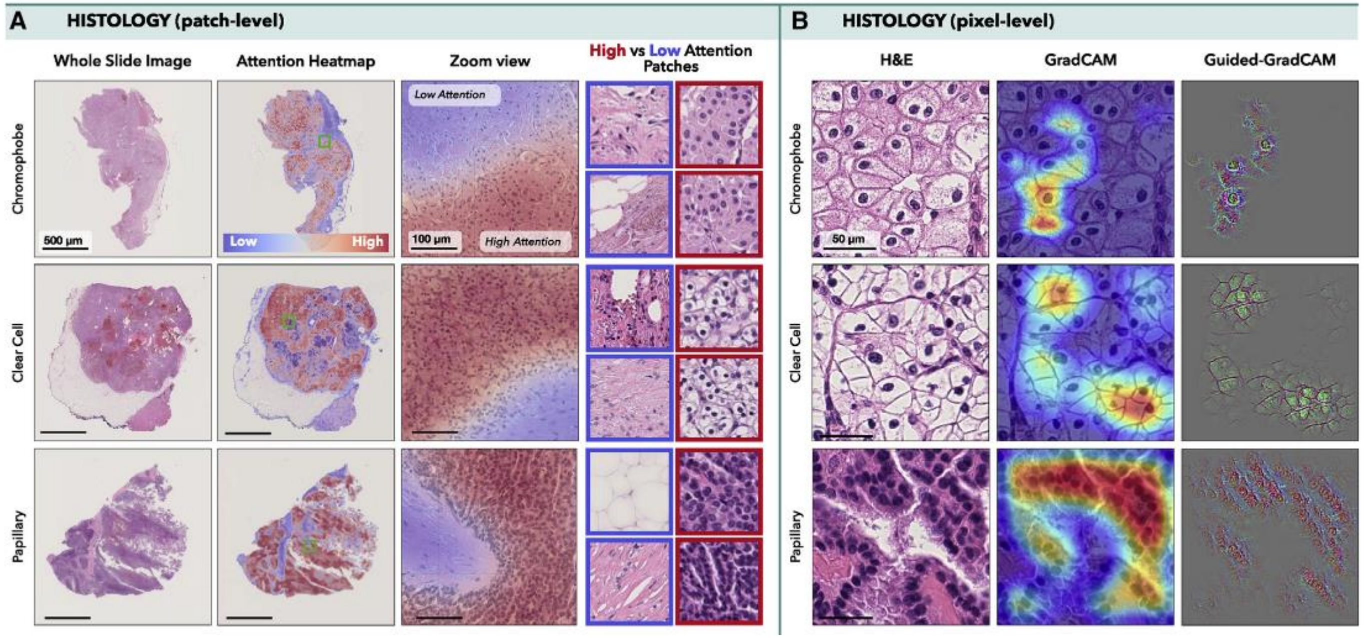
Selvaraju , et al. (2017). Grad-cam: visual explanations from deep networks via gradient-based localization. In ICCV, pp. 618–626.
Chattopadhay , et al. (2018). Grad-cam++: generalized gradient- based visual explanations for deep convolutional networks. In WACV, pp. 8
Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Gradcam, D.B. (2016). Why did you say that?. Preprint at arXiv, 1611.07450.

Interpretability

Histopathology

Depending on the model architecture **attention or probability scores** can be mapped to obtain **slide-level** attention heatmaps.

- In the **attention** methods, the importance of each instance is determined **inside the model**.
- The **CAM-based** methods are **model agnostic**, i.e., independent of the models.



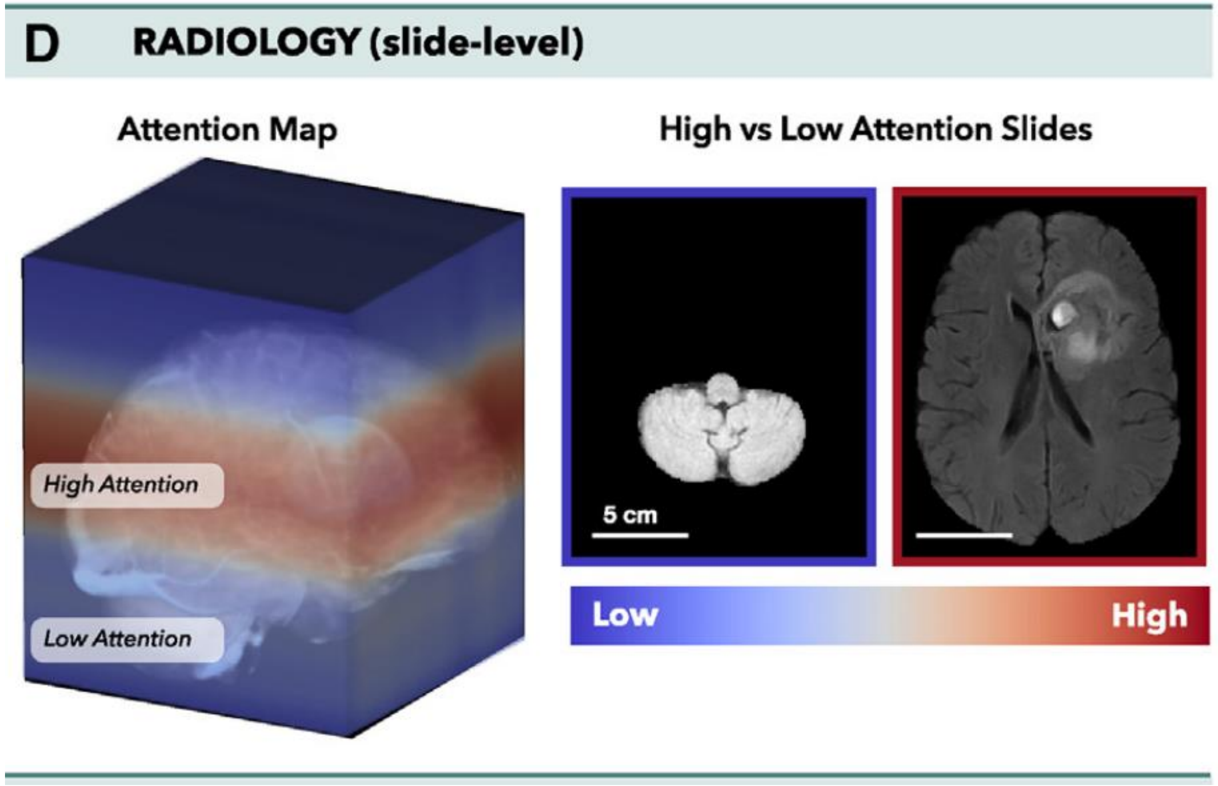
Selvaraju , et al. (2017). Grad-cam: visual explanations from deep networks via gradient-based localization. In ICCV, pp. 618–626.
Chattopadhay , et al. (2018). Grad-cam++: generalized gradient- based visual explanations for deep convolutional networks. In WACV, pp. 8
Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., and Gradcam, D.B. (2016). Why did you say that?. Preprint at arXiv, 1611.07450.

Interpretability

Radiology

- Similar to those used in histology.

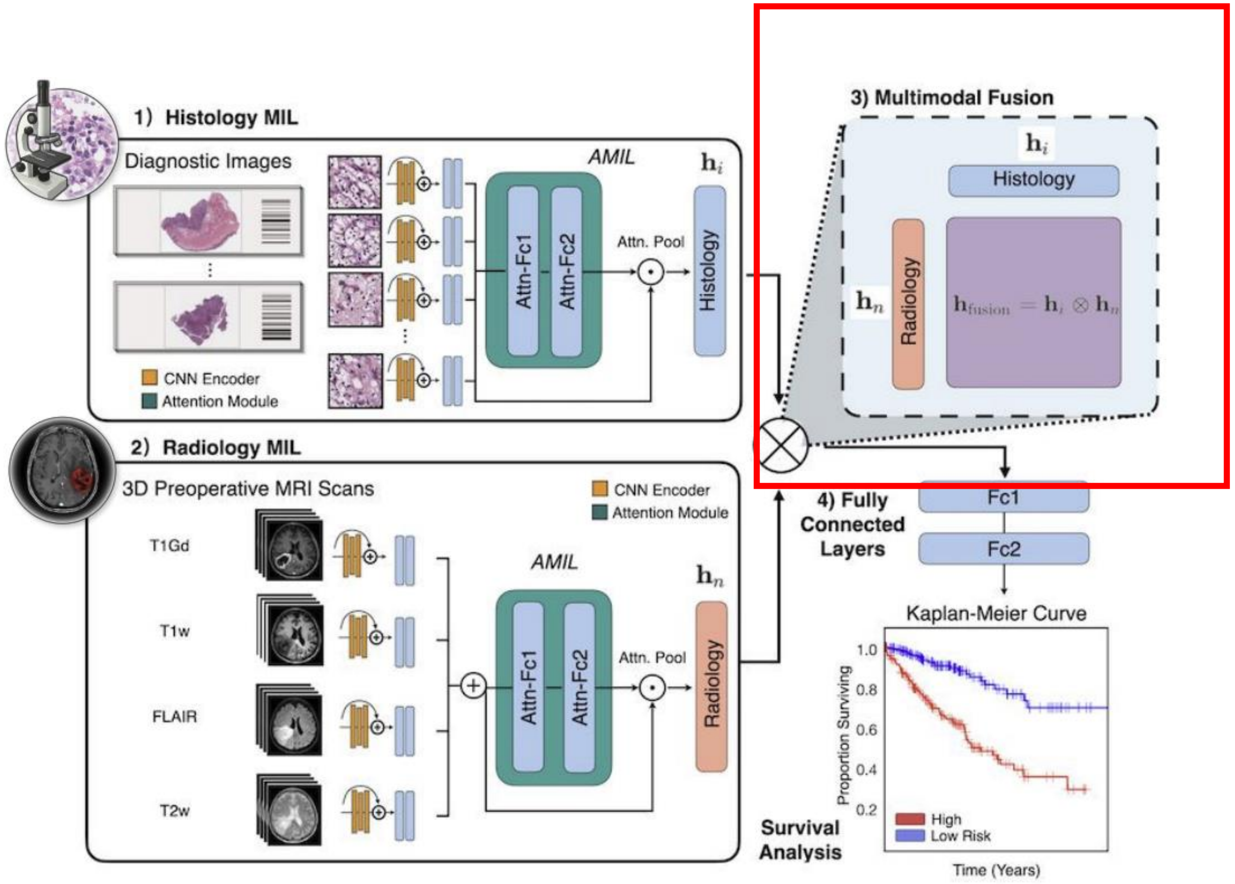
- The **attention** scores can reflect the **importance of slides** in a 3D scan.



Interpretability

Radiology

- The model considered the 3D MRI **scan** as a **bag**, where the axial **slices** are modeled as individual **instances**.
- The model placed **high attention** to the slices with **tumor**, while **low attention** was assigned to **healthy tissue**.

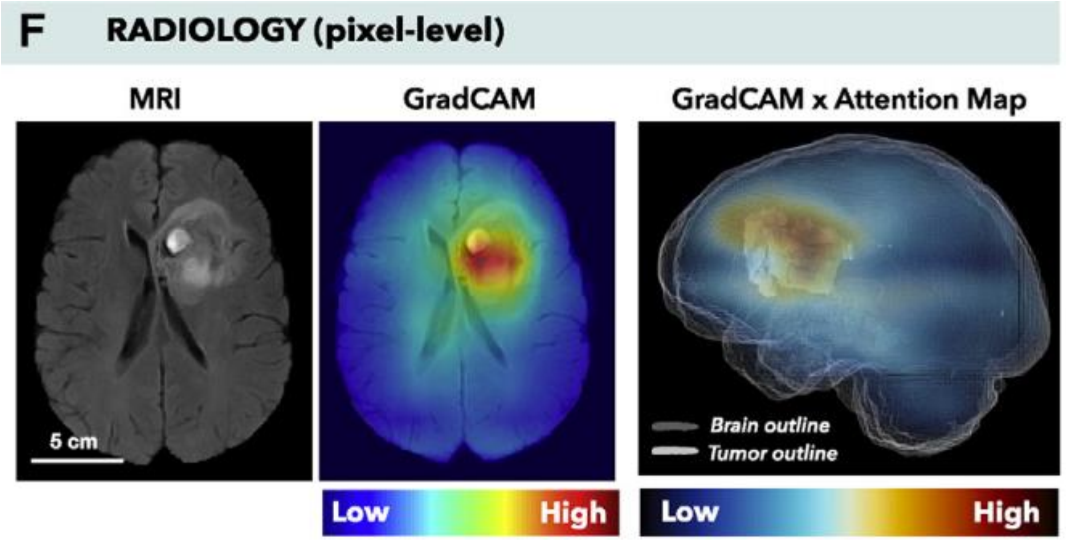


Zhuang, L., Lipkova, J., Chen, R., and Mahmood, F. (2022). Deep learning-based integration of histology, radiology, and genomics for improved survival prediction in glioma patients. In In Medical Imaging 2022: Digital and Computational Pathology, 12039 (SPIE), p. 120390Z.

Interpretability

Radiology

- CAM-based methods can be deployed to **localize the predictive regions** within individual slices.



Zhuang, L., Lipkova, J., Chen, R., and Mahmood, F. (2022). Deep learning-based integration of histology, radiology, and genomics for improved survival prediction in glioma patients. In In Medical Imaging 2022: Digital and Computational Pathology, 12039 (SPIE), p. 120390Z.

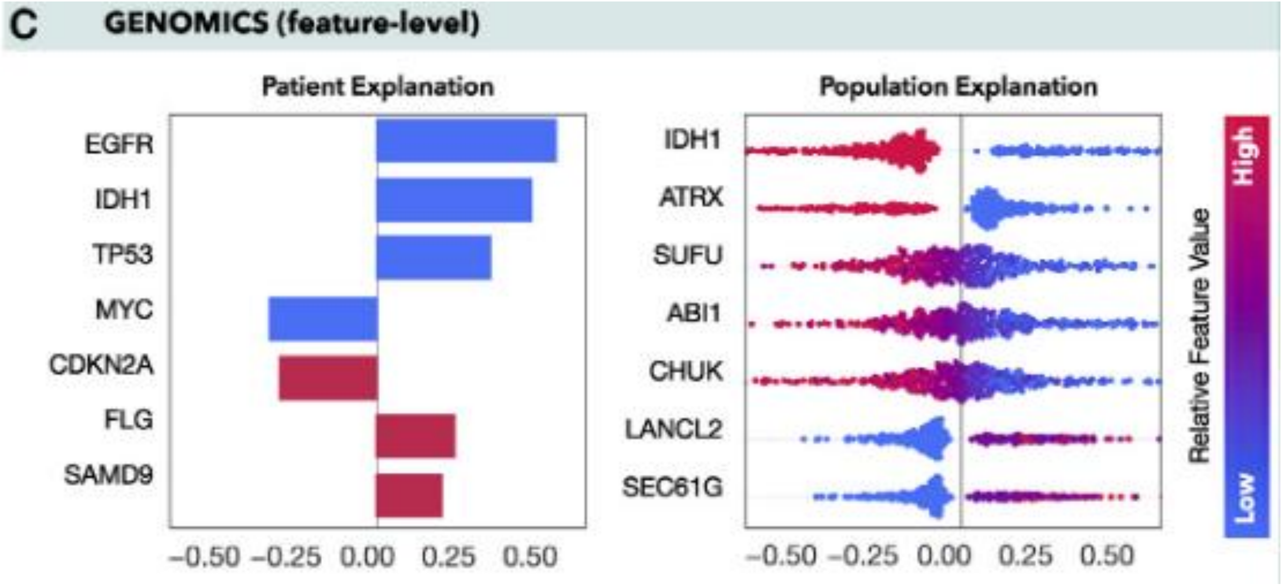
Interpretability

Molecular data

- Compute **attribution values** indicating how changes in specific inputs affect the model outputs.

For example, **in survival analysis**, the attribution values can reflect the **magnitude of the importance** as well as the **direction of the impact**:

- Features with **positive** attribution increase the predicted output (i.e., **higher risk**).
- Features with **negative** attribution reduce the predictive values (i.e., **lower risk**).

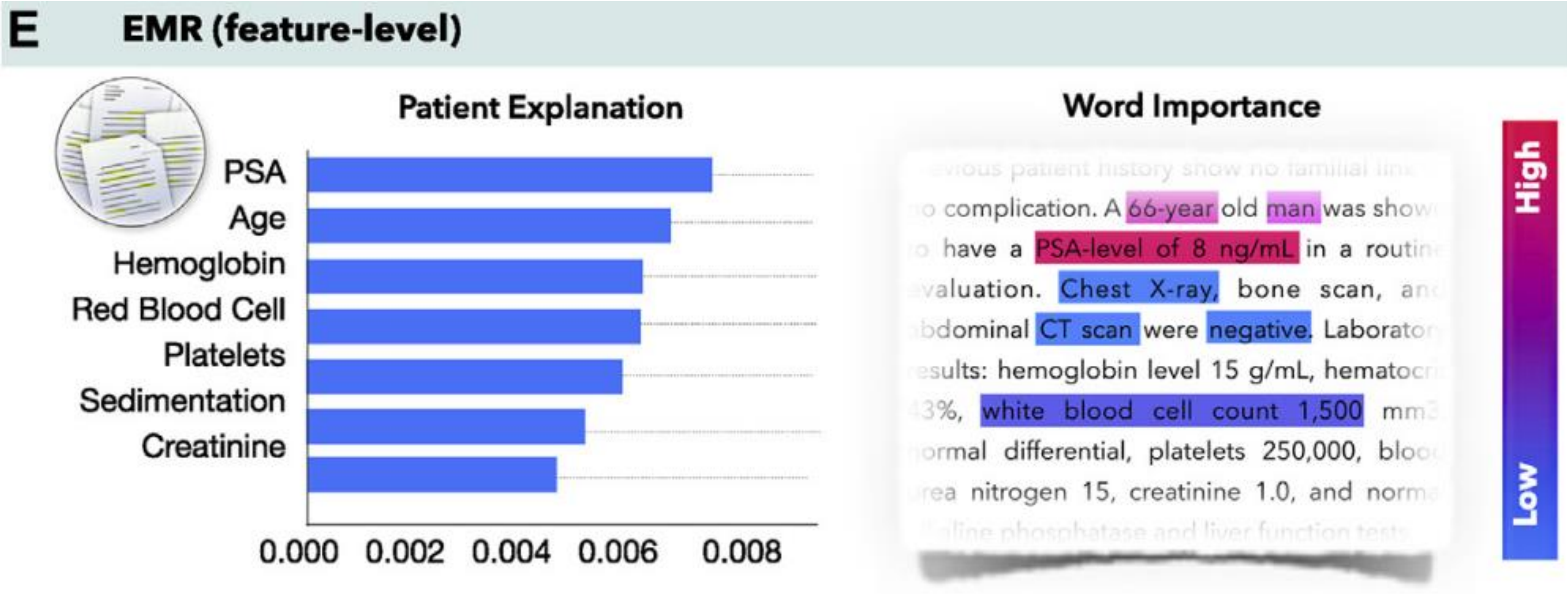


Chen R J, Lu M Y, Williamson D F K, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning[J]. Cancer Cell, 2022, 40(8): 865-878. e6.

Interpretability

Molecular data

Electronic Medical Records (EMRs) can be also analyzed by natural language processing (NLP) methods, such as transformers, where the **attention scores** determine the importance of **specific words** in the text.



Interpretability

Summary

- All previously mentioned methods can be used in multimodal models to explore interpretability within each modality.
- The interpretability methods usually come **without any quantitative measures**, and thus it is important not to **overinterpret** them.
- They can explain ***where*** but not ***why***.
  - e.g., CAM- or attention-based methods can localize the predictive regions, they cannot specify which features are relevant.
  - There is **no guarantee** that all high-attention/attribution regions carry clinical relevance.

Multimodal Learning in Healthcare

- Introduction
- Methods
- Interpretability
- Multimodal data interconnection
- Challenge and clinical adoption

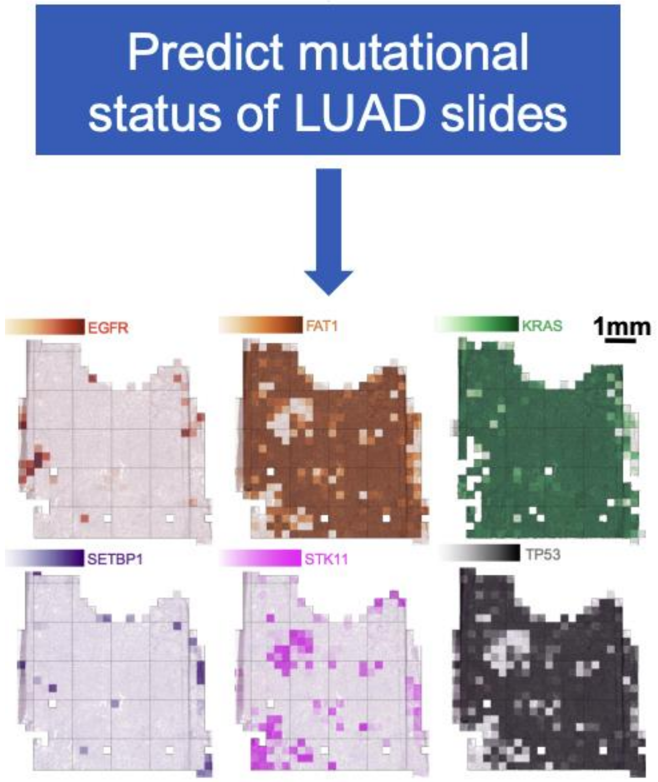
Multimodal data interconnection

- The aim of multimodal data interconnection is to **reveal associations and shared information** across modalities.
- Such associations can **provide new insights** into cancer biology and **guide the discovery of novel biomarkers**.

Multimodal data interconnection

Morphologic associations

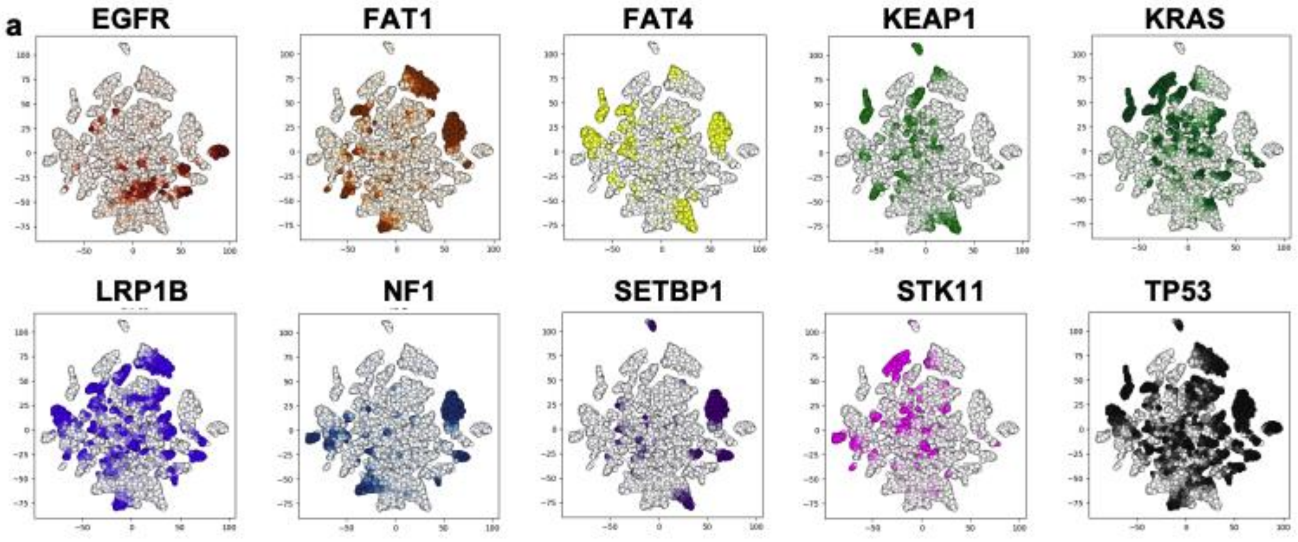
- Malignant changes often propagate across **different scales**.
  - **Oncogenic mutations** can affect cell behavior, which in turn reshapes **tissue morphology or the tumor microenvironment** visible in histology images.



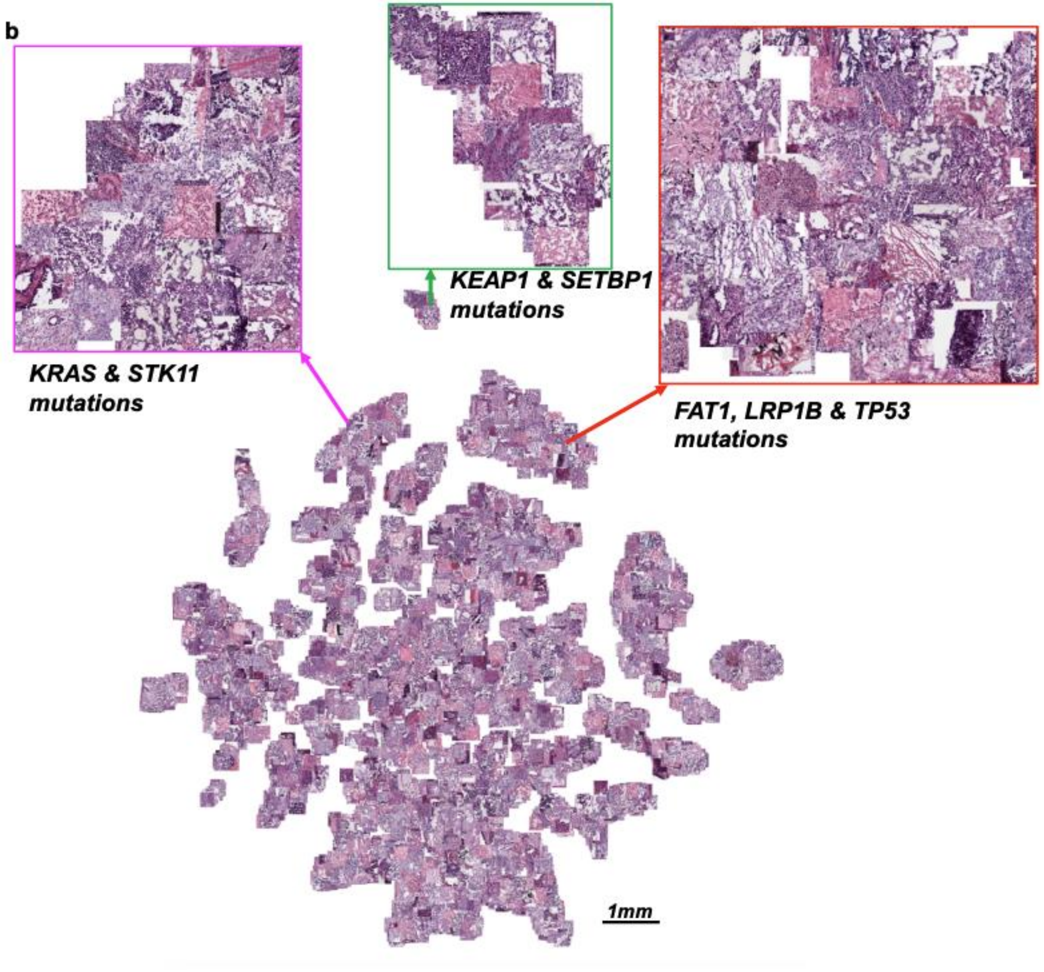
Coudray N, Ocampo P S, Sakellaropoulos T, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning[J]. Nature medicine, 2018, 24(10): 1559-1567.

Multimodal data interconnection

Morphologic associations - microscopic



- Each dot represents a tile, and its color is proportional to the probability of the gene to be mutated.



- Certain **mutations** can be **inferred directly** from hematoxylin and eosin (H&E)-stained WSIs.

Coudray N, Ocampo P S, Sakellaropoulos T, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning[J]. Nature medicine, 2018, 24(10): 1559-1567.

Multimodal data interconnection

Morphologic associations – macroscopic

- Predicted IDH mutation and 1p/19q codeletion status from preoperative brain MRI scans.

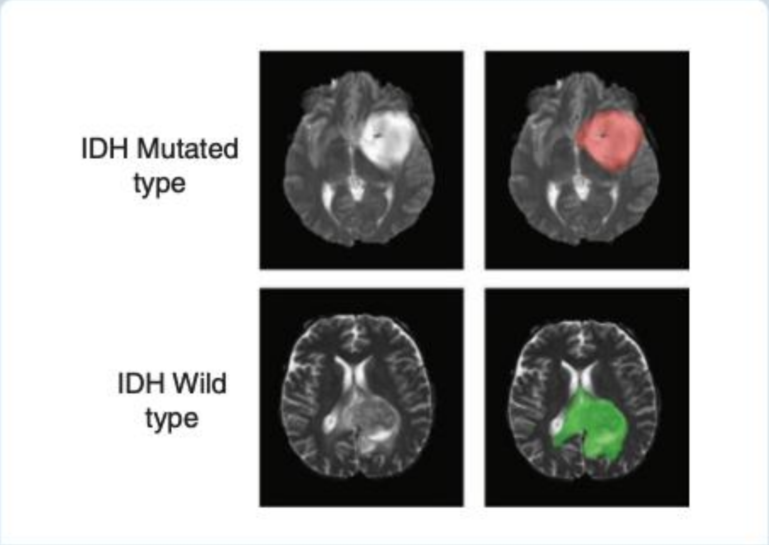


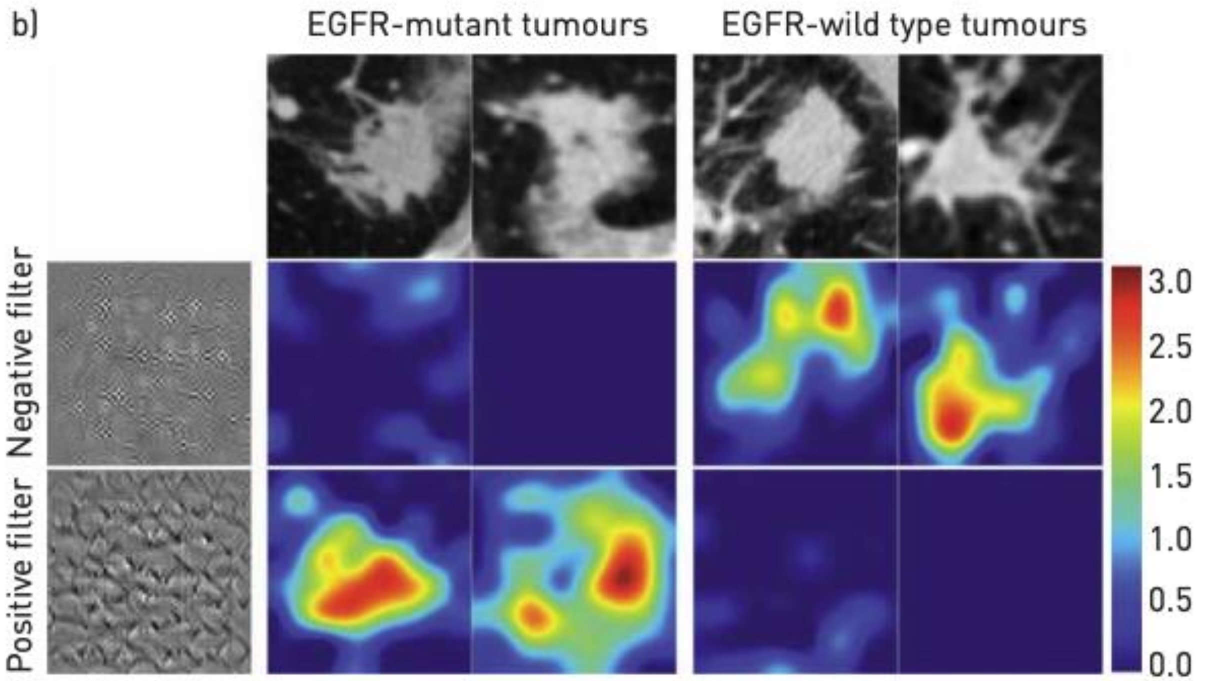
Fig. 1 Ground truth whole tumor masks. Red voxels represent IDH mutated (value of 1) and green voxels represent IDH wild-type (value of 2). The ground truth labels have the same mutation status for all voxels in each tumor.

Bangalore Yogananda C G, Shah B R, Vejdani-Jahromi M, et al. A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas[J]. Neuro-oncology, 2020, 22(3): 402-411.

Multimodal data interconnection

Morphologic associations – macroscopic

- EGFR and KRAS mutations have been detected from CT scans in lung.

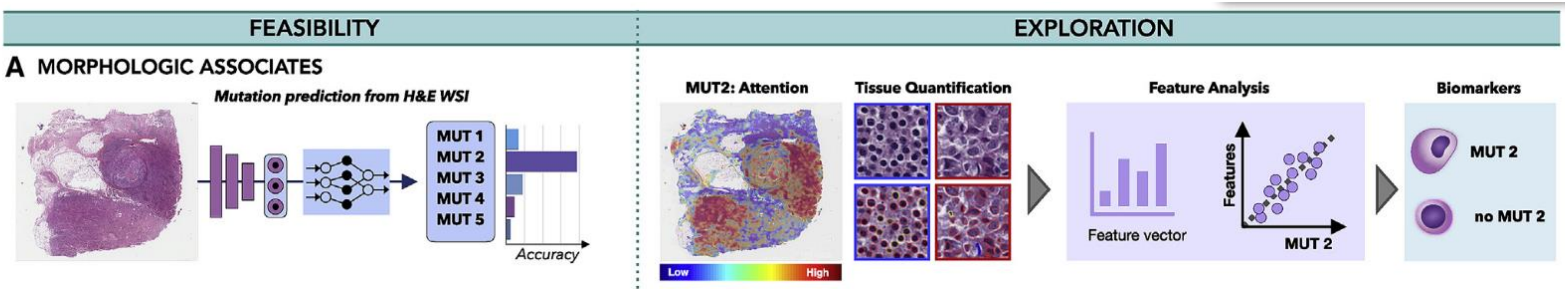


Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning[J]. European Respiratory Journal, 2019, 53(3).

Multimodal data interconnection

Morphologic associations

- By discovering the presence of **morphological associations** across modalities, AI models can **enhance exploratory studies and reduce the search space** for **possible biomarker** candidates.



AI has revealed that one of the studied mutations can be reliably inferred from WSI.

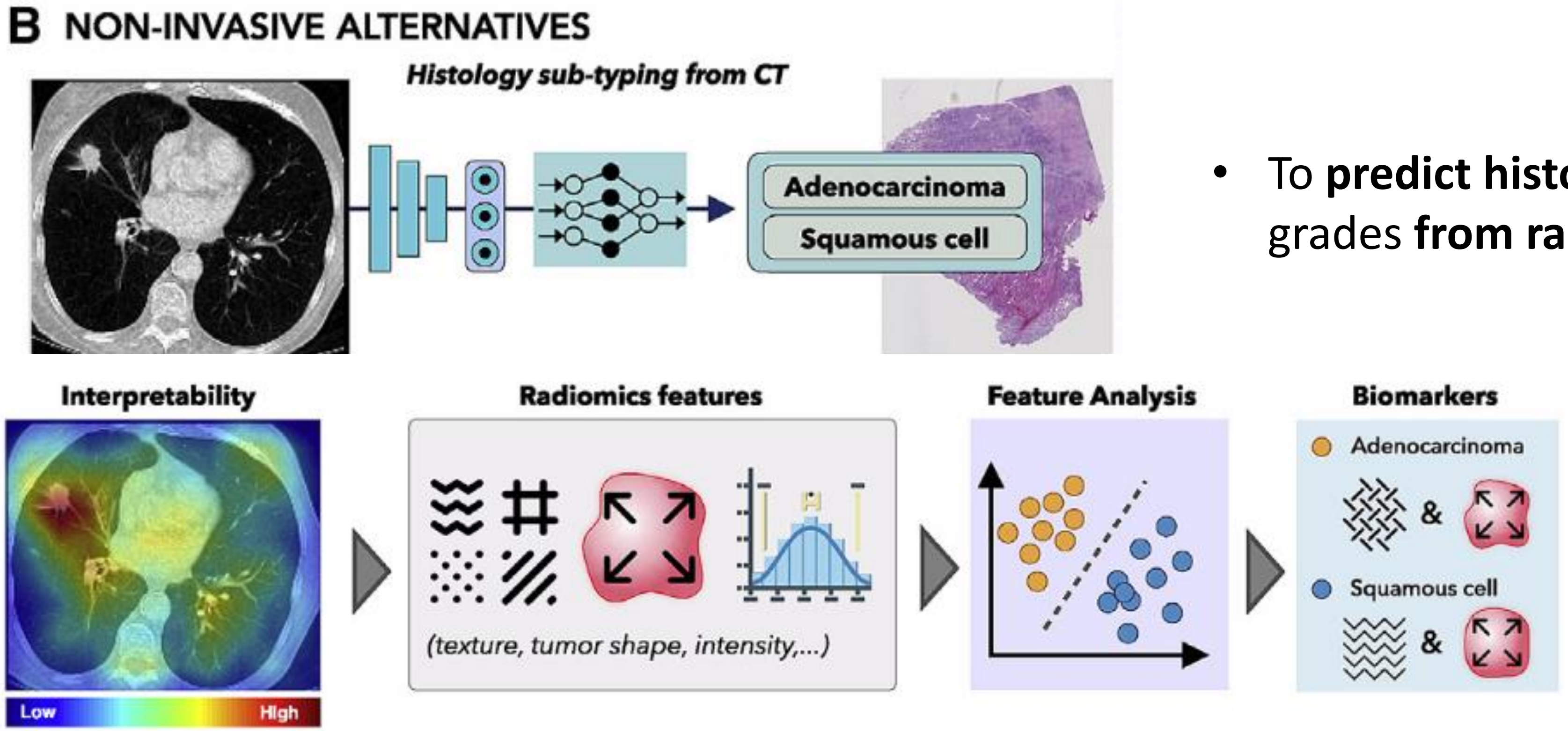
Interpretability methods

- The identified morphological associates can serve as **cost-efficient biomarker surrogates.**

Multimodal data interconnection

Associations between **non-invasive** and **invasive modalities**

- Serve as **non-invasive surrogates** for existing biomarkers of invasive modalities.



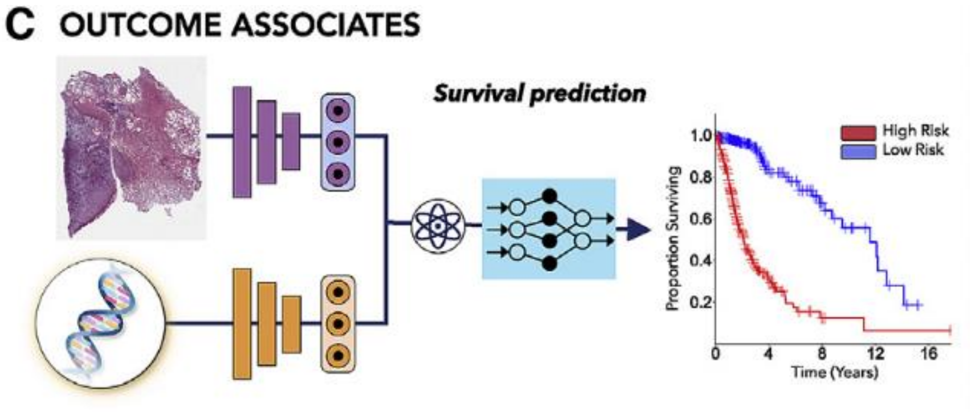
- To **predict histology** subtypes or grades **from radiomics features**.
- The predictive image regions can be further analyzed to **identify textures and patterns** with **possible diagnostic value**.

Sha X, Gong G, Qiu Q, et al. Identifying pathological subtypes of non-small-cell lung cancer by using the radiomic features of 18F-fluorodeoxyglucose positron emission computed tomography[J]. Translational Cancer Research, 2019, 8(5): 1741.

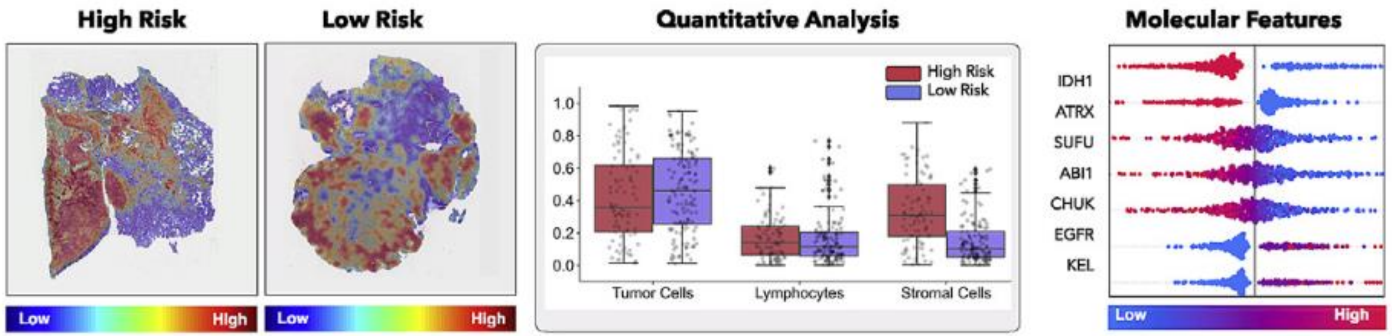
Multimodal data interconnection

Associations with Outcome

- The feasibility of AI models to discover **relevant prognostic patterns** in data.



- Attention heatmaps reveal tissue regions related to **low- and high-risk patient groups**.
- The molecular profiles are analyzed through **attribution plots**.

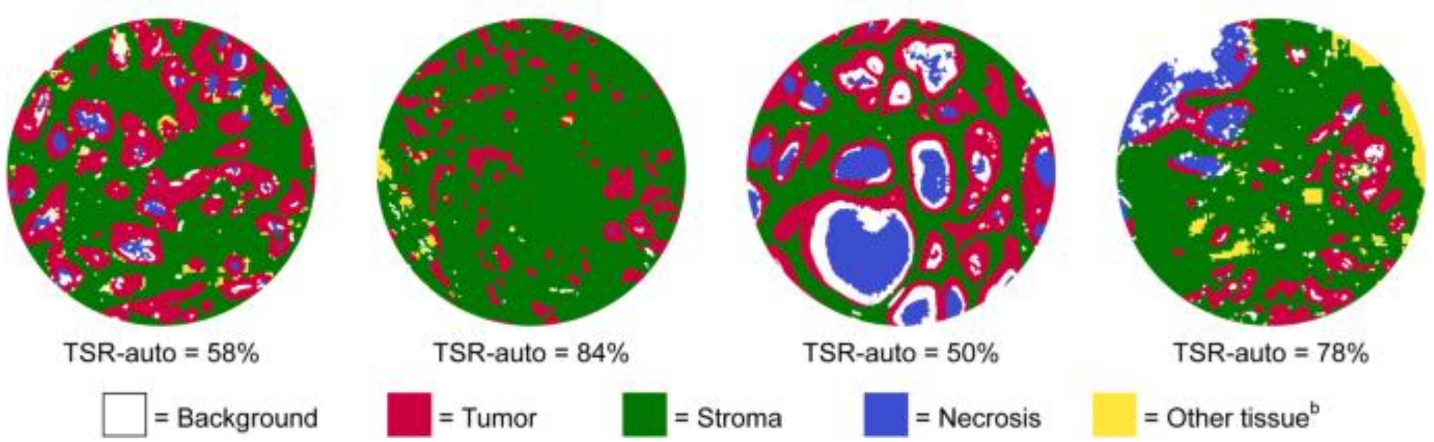


- Such exploration studies have already **provided new clinical insights**.

Multimodal data interconnection

Associations with Outcome

- The **tumor-to-stroma ratio** can serve as an **independent prognosticator** in rectal cancer.



A deep learning algorithm for **automated** TSR assessment (TSR-auto).

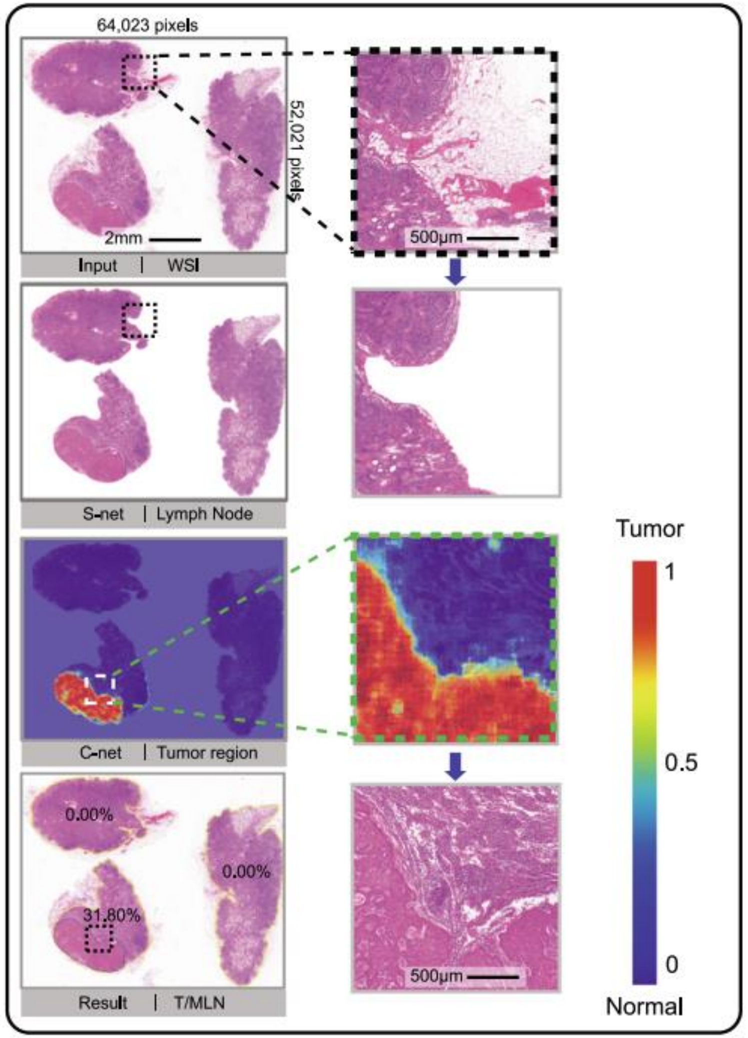
Geessink O G F, Baidoshvili A, Klaase J M, et al. Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer[J]. Cellular oncology, 2019, 42: 331-341.

Multimodal data interconnection

Associations with Outcome

- The **ratio of tumor area to metastatic lymph node regions** has prognostic value in gastric cancer.

$$\text{T/MLN} = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{A^{i}_{\text{tumor}}}{A^{i}_{\text{MLN}}}\right)$$

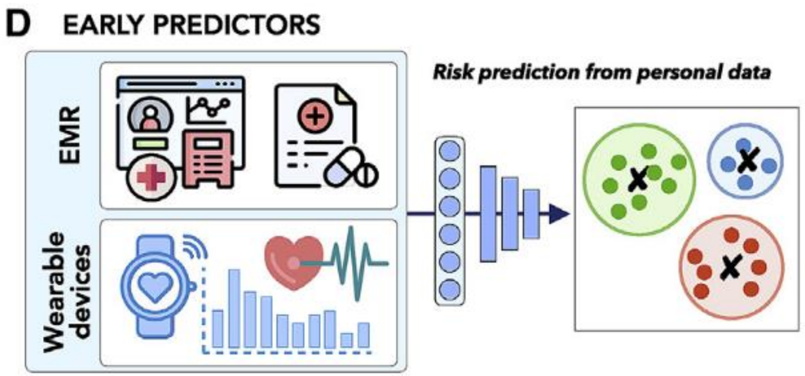


Wang X, Chen Y, Gao Y, et al. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning[J]. Nature communications, 2021, 12(1): 1637.

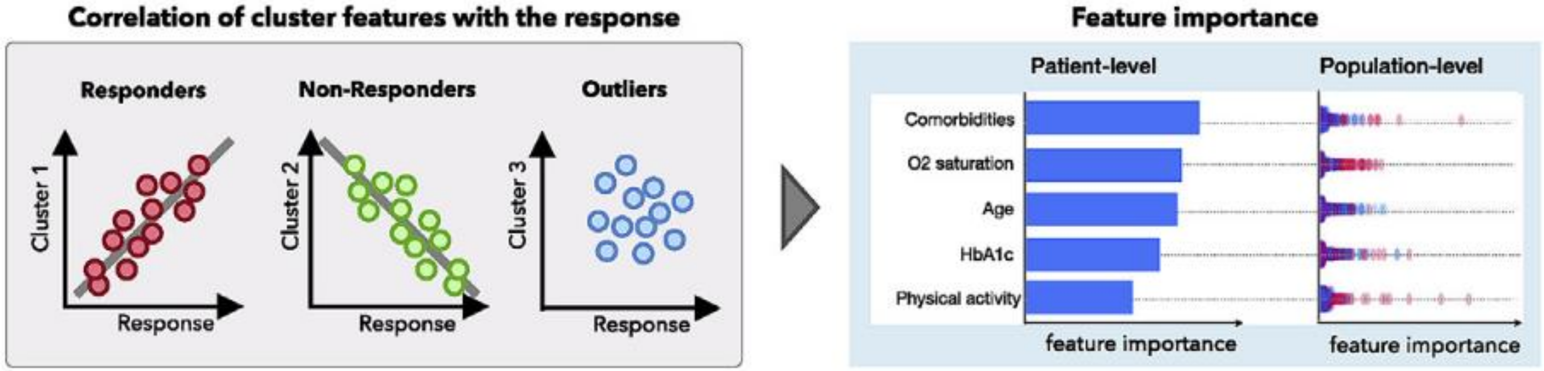
Multimodal data interconnection

Associations with Early Predictors

- AI can also explore various modalities acquired prior to patient diagnosis to identify **potential predictive risk factors**.



- **EMRs** provide rich information on patient history, medication, allergies, or immunizations.
- Information acquired by **EMRs** or **wearable devices** can be analyzed to identify risk factors to support **early interventions**.



- To search for distinct patient subgroups.

Multimodal Learning in Healthcare

- Introduction
- Methods
- Interpretability
- Multimodal data interconnection
- Challenge and clinical adoption

Challenge and clinical adoption

The path of AI into clinical practice is still laden with obstacles, many of which are amplified in the presence of multimodal data.

Challenges		
Missing data	Data alignment	Transparency and prospective clinical trials

Challenge and clinical adoption

Missing data

Possible strategies:

1. Synthetic data generation
   - If **part** of an image is corrupted, or if specific mutations are not reported, the missing information can be synthesized from the remaining data.
   - If **a whole modality** is missing, its synthetic version can be derived from existing similar modalities.
2. Dropout-based methods
   - Dropout-based methods aim to make models **robust to missing information**.

Challenge and clinical adoption

Data alignment

To investigate cancer processes across different **scales and modalities**, a certain level of data alignment is required.

This might include alignment of (1) **diverse** or (2) **similar** modalities.

Alignment of diverse modalities

- This refers to the integration of data from different **scales, time points, or measurements**.
- Often an acquisition of one modality results in the destruction of the sample, preventing collection of multiple measurements from the same system.

Challenge and clinical adoption

Data alignment

Alignment of similar modalities

- This method typically involves different imaging modalities of the **same system**.
- This is usually achieved through **image registration**, which is formulated as an optimization problem minimizing the difference between the modalities.

For examples,

- In histology, each **stained slide** usually comes from a different tissue cut.
- In radiology, **rigid anatomical structures** can guide the data alignment.

Challenge and clinical adoption

Transparency and prospective clinical trials

- **Abstract representation** learning-based modern AI methods, **cannot be fully understood**, although some of the interpretability methods can indicate predictive regions.
- We should advocate for their **rigorous validation under randomized clinical trials**, same as is done for other medical devices and drugs.
- **Prospective clinical trials** are inevitable to truly demonstrate and quantify the added value of AI models, which will in turn **increase trust and motivation** of practitioners toward the clinical adoption of AI tools.