Towards a multimodal AI resilient to data in-the-wild

Massimo Esposito[1], **Michela Gravina[2]**, Valerio Guarrasi[3], Giulio Iannello[3], Carlo Sansone[2], Paolo Soda[3], Mario Vento[1]

[1]Institute for High-Performance Computing and Networking (ICAR) - Research National Council of Italy (CNR), Italy
[2]Department of Electrical Engineering and Information Technology, University of Naples Federico II, Napoli
[3]Department of Engineering, University of Rome Campus Bio-Medico, Roma

Naples, 23rd September 2024

GENERAL CONFERENCE
FAIR Future Artificial Intelligence Research
23|24SET2024
NAPOLI

Multimodal Artificial Intelligence

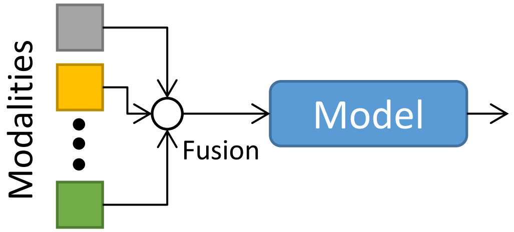
- Multimodal Learning allows the fusion of complementary information coming from heterogeneous sources

Early Fusion

The integration of heterogeneous sources of data in a single structure that is then used as input for a model

✓ It only requires a single learning phase

X It is challenging to combine all the modalities into a common representation

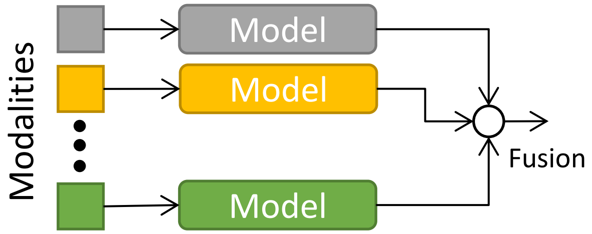


Late Fusion

The integration of the decision coming from different models, each trained on a specific image modality

✓ It exploits the unrelated errors of different models

X It has a high computation cost

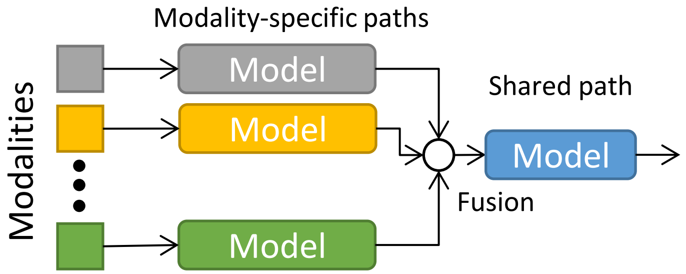


Intermediate Fusion

The integration of units coming from multiple modality-specific paths in a single representation, known as shared representation

✓ It is a flexible method, allowing for data fusion at different stages of model training

X It is challenging to choose which modality to fuse at which depth of representation



Resilience in Multimodal AI

1 Modality Integration Challenges

2 Dealing with Noisy Data

3 Balancing Multimodal Inputs

4 Computational Resource

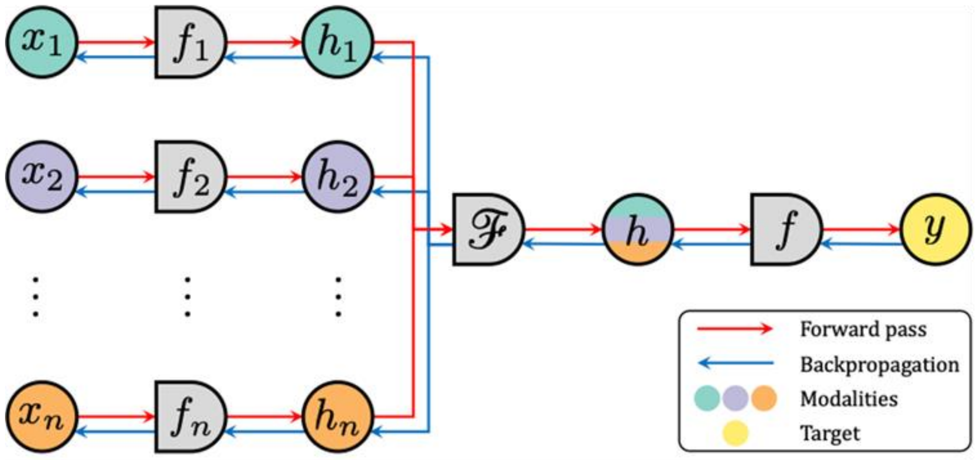
5 Ethical and Bias Challenges

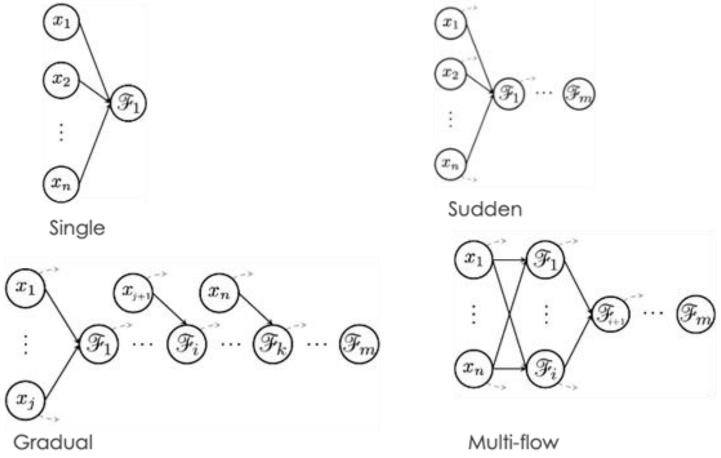
6 Transparency

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

1 Modality Integration Challenges: Intermediate Fusion Analysis

- Systematic Review of Intermediate Fusion Methods of Multimodal Deep Learning in Biomedical Applications:
  - **Comprehensive Methodology for Taxonomy Development:** We have devised a robust methodology to create a taxonomy that assists in selecting appropriate model configurations tailored to specific modalities and tasks
  - **Enhanced Decision-Making Framework:** Our approach facilitates informed decision-making by guiding the selection of optimal configurations for various model components in multimodal scenarios



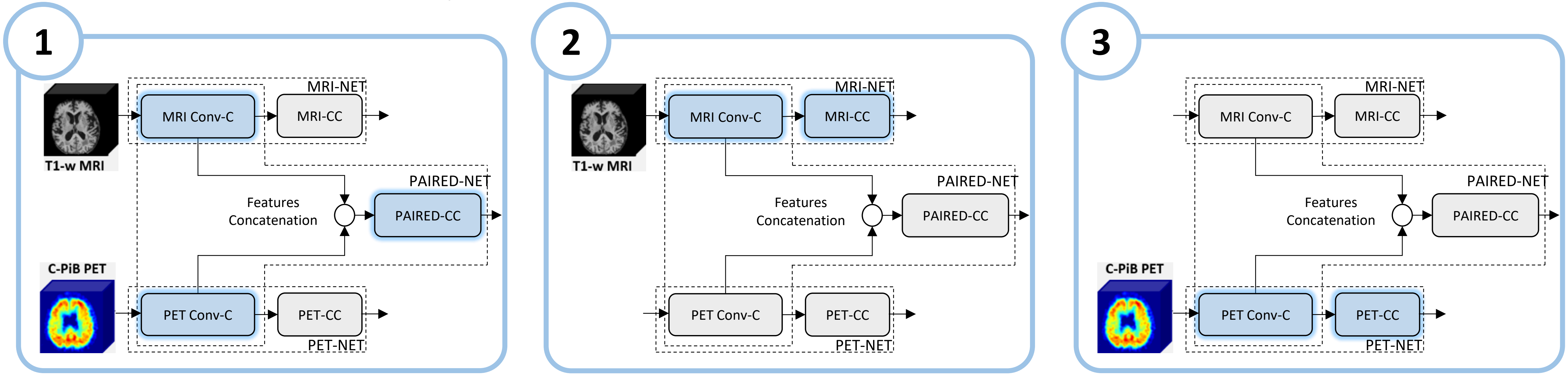


Link:

Guarrasi, V., Aksu, F., Caruso, C. M., Di Feola, F., Rofena, A., Ruffini, F., & Soda, P. (2024). A Systematic Review of Intermediate Fusion in Multimodal Deep Learning for Biomedical Applications. arXiv preprint arXiv:2408.02686.

2 Dealing with Noisy Data: Fusion strategies with Incomplete Dataset

- Strategy to handle the challenges posed by and unbalanced dataset with incomplete acquisitions
  - As case of study, dementia severity assessment in MRI and PET is considered
- A Multi Input – Multi Output 3D Convolutional Neural Network is used to process paired and incomplete data with an innovative training strategy that makes the network able to change the training step according to the characteristics of the input

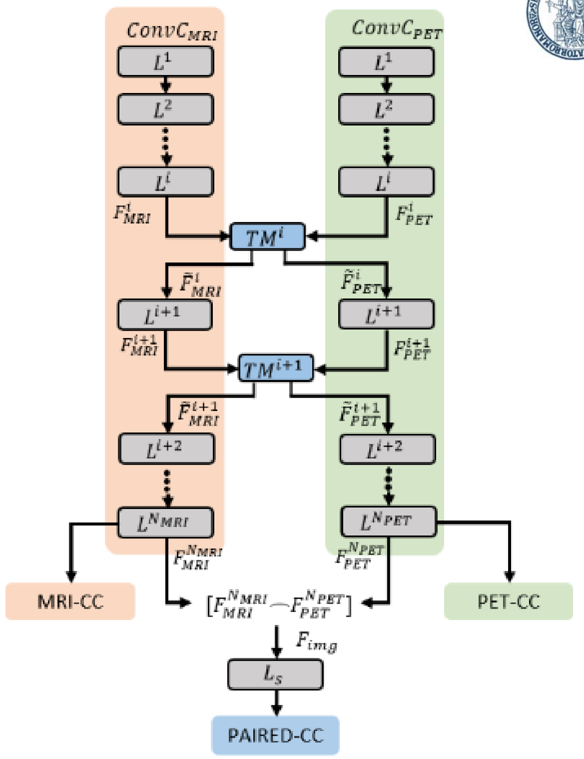


Gravina, M., García-Pedrero, A., Gonzalo-Martín, C., Sansone, C., & Soda, P. (2024). Multi input–Multi output 3D CNN for dementia severity assessment with incomplete multimodal data. *Artificial Intelligence in Medicine*, *149*, 102774.

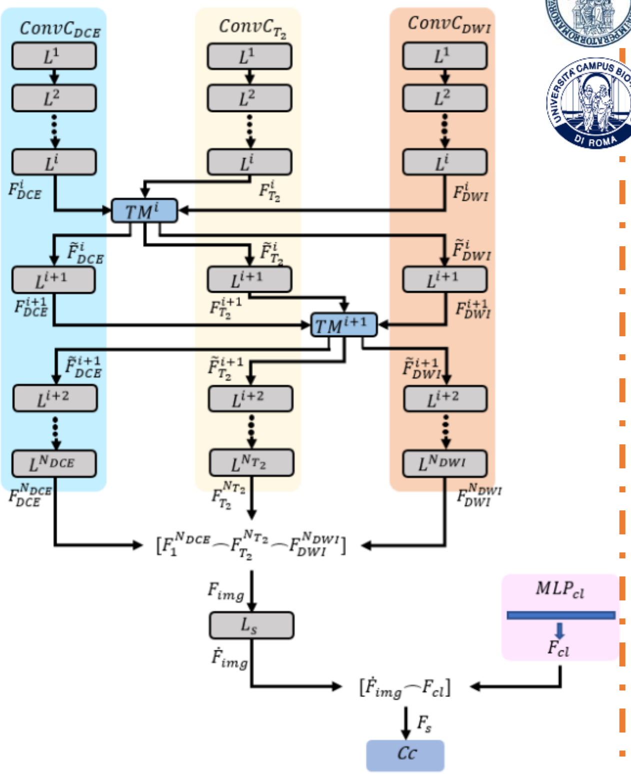
Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

3 Balancing Multimodal Inputs: Cross-modality learning techniques

- **Transfer module (TM)**: a module that improves the integration of heterogeneous and multimodal sources of data in the intermediate fusion technique. The transfer module (TM) performs a cross-modality calibration of the extracted features highlighting the effects of the most discriminative characteristics while reducing the less useful ones
- We consider two different scenarios:
  - **Paired Dataset**: axillary lymph node metastasis prediction in breast cancer using multiparametric MRI and histological data
  - **Incomplete Dataset**: dementia severity assessment in MRI and PET



Incomplete Dataset

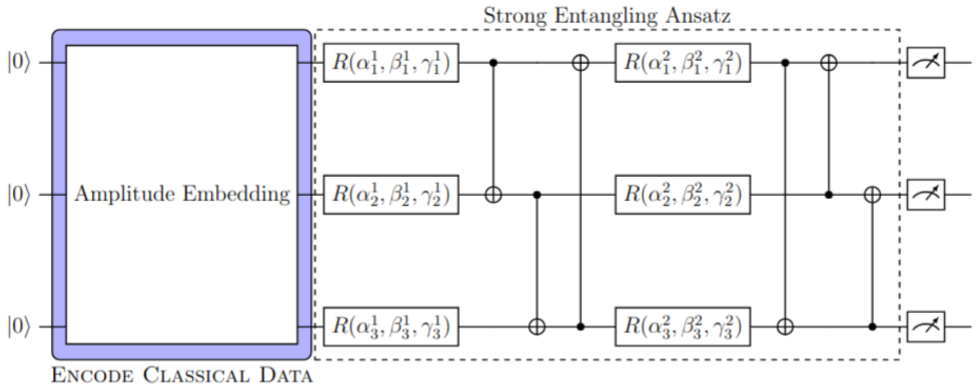


Paired Dataset

- Gravina, M., Santucci, D., Cordelli, E., Soda, P., Sansone C., Cross-Modality Calibration in Multi-Input Network for Axillary Lymph Node Metastasis Evaluation. IEEE Transactions on Artificial Intelligence, 2024.
- De Simone, A., Gravina M., Sansone C., Multimodality Calibration in 3D Multi Input-Multi Output Network for Dementia Diagnosis with Incomplete Acquisitions, accepted to IAPR International Workshops on Statistical Techniques in Pattern Recognition and Structural and Syntactic Pattern Recognition (S+SSPR)

4 Computational Resource: Multi-task Multi-language Quantum Transfer Learning

- Modern LLMs needs a huge amount of data for their training to be effective
- They might struggle with ambiguity and polysemy intrinsic in any language
- For various NLPs tasks, models need to be fine-tuned and re-built to accommodate the proper solution



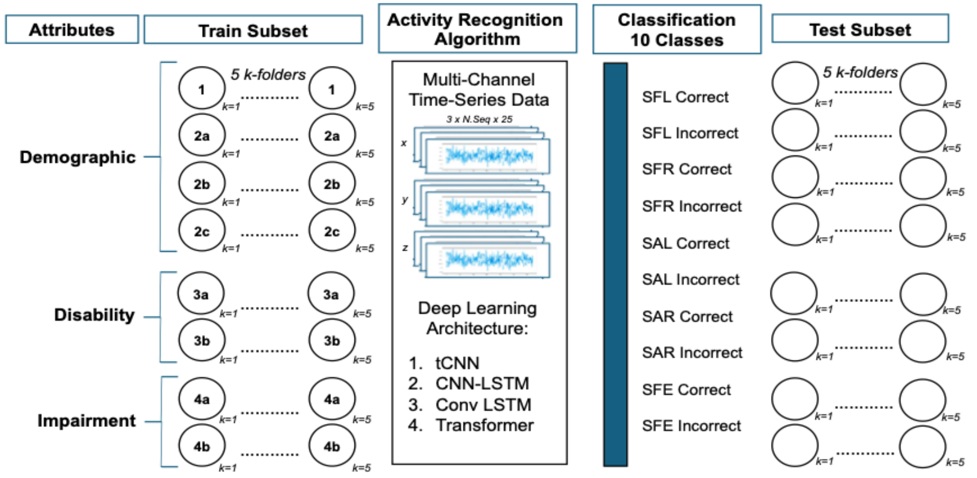
Dataset	Bert Classical	Bert Quantum	Electra Classical	Electra Quantum
Cola	0.815 ±0.008	0.795 ±0.008	**0.842** ±0.005	**0.842** ±0.005
ItaCola	0.904 ±0.05	0.899 ±0.009	**0.923** ±0.008	0.920 ±0.008
STN-2	0.910 ±0.005	0.920 ±0.008	0.942 ±0.006	**0.945** ±0.008
SentiPolc	0.755 ±0.006	0.760 ±0.008	0.755 ±0.005	**0.770** ±0.005

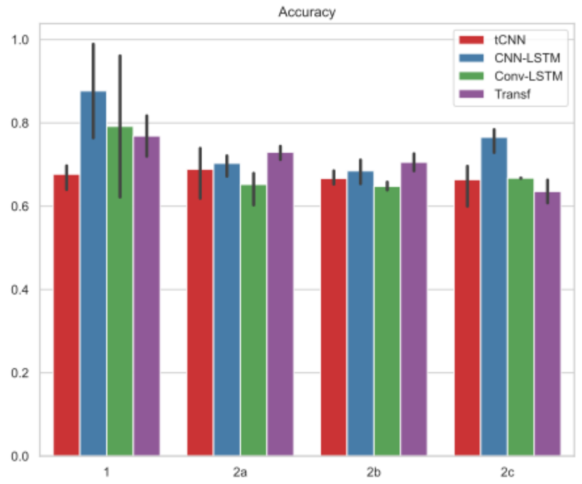
- Quantum Computing allows to represent embedded knowledge- coming from classical pre-trained LLMS- in a large vector space and to extract richer information compared to classical models
- Quantum Transfer Learning can outperforms Standard Classifiers, furthermore allows to distinguish sub-types of linguistic structures, thus its proven to be adequate to tackle cross-languages multiple tasks in NLP

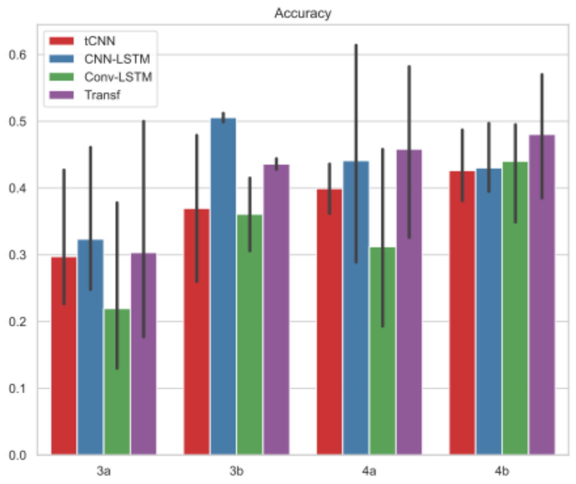
- Buonaiuto G., Guarasci R, De Pietro G., Esposito M. *Quantum Transfer Learnin for Acceptability Judgement*. Quantum Machine Intelligence 6, 13, 2024.
- Buonaiuto G., Guarasci R, De Pietro G., Esposito M. *Multitask Multilingual quantum transfer learning,*. currently under review

5 Ethical and Bias Challenges: Promoting Fairness in Activity Recognition Algorithm

- The research examines how **intrinsic attributes of subjects** can influence **model fairness** in **healthcare monitoring and support systems**, and explores potential solutions to address these biases
- **Kinematic data** (x, y, z coordinates of 25 joints) of **28 subjects** performing different rehabilitation activities
- Deep learning models **processing multi-channel time series data** (tCNN, hybrid-LSTM, Transformer)
- **Different balancing settings** for **demographic, disability, and physical impairment attributes** in training subsets to assess their impact on **algorithm bias**



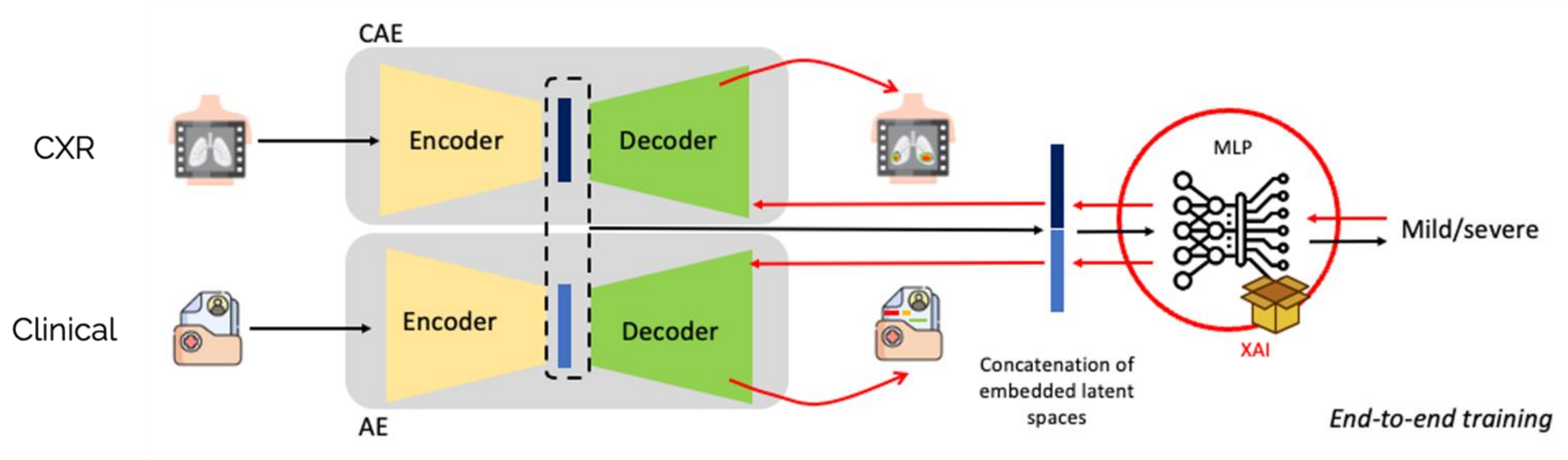




Mennella, C., Esposito, M., De Pietro, G., & Maniscalco, U. (2024). Promoting fairness in activity recognition algorithms for patient's monitoring and evaluation systems in healthcare. Computers in biology and medicine, 179, 108826.

6 Transparency: Multimodal XAI

- The major disadvantage of DNNs is their **lack of interpretability**: XAI produces information to make a model's functioning clear or easy to understand. The literature is well advanced for unimodal models but it lacks research for Multimodal Deep Learning
- **Method:**
  - Multi-modal loss and three-stage training
  - Counterfactual explanations
  - It is possible to show the relative contribution of each modality in making the decision



Guarrasi, V., Tronchin, L., Albano, D., Faiella, E., Fazzini, D., Santucci, D., & Soda, P. (2024). Multimodal explainability via latent shift applied to COVID-19 stratification. Pattern Recognition, 156, 110825.

Conclusions

- Resilient multimodal AI is essential for deploying systems in real-world environments, ensuring robustness, adaptability, and reliability in complex, unpredictable conditions

- Key Aspects of Multimodal AI:
  - Handling Real-World Complexity
  - Robustness to Noisy Data
  - Adaptability to Dynamic Environments
  - Cross-modal Compensation
  - Enhanced Contextual Understanding

- Future works:
  - Evaluating the generalization ability of the proposed methods for different scenarios and applications
  - Analysis of computational cost of multimodal systems
  - Enhancing robustness to adversarial attacks

Thank you for your Attention!

GENERAL CONFERENCE
FAIR Future Artificial Intelligence Research
23|24SET2024
NAPOLI