

# Towards AI-driven Next Generation Personalized Healthcare and Well-being

Fatih Aksu<sup>1</sup>, Alessandro Bria<sup>2</sup>, Alice Natalina Caragliano<sup>3</sup>, Camillo Maria Caruso<sup>3</sup>, Wenting Chen<sup>4</sup>, Ermanno Cordelli<sup>3,\*</sup>, Omar Coser<sup>3,5</sup>, Arianna Francesconi<sup>3</sup>, Leonardo Furia<sup>3</sup>, Valerio Guarrasi<sup>3</sup>, Giulio Iannello<sup>3</sup>, Clemente Lauretti<sup>5</sup>, Guido Manni<sup>3,5</sup>, Giustino Marino<sup>3</sup>, Domenico Paolo<sup>3</sup>, Filippo Ruffini<sup>3</sup>, Linlin Shen<sup>6</sup>, Rosa Sicilia<sup>3</sup>, Paolo Soda<sup>3,7</sup>, Christian Tamantini<sup>5</sup>, Matteo Tortora<sup>3</sup>, Zhuoru Wu<sup>6</sup> and Loredana Zollo<sup>5</sup>

<sup>1</sup>Department of Biomedical Sciences, Humanitas University, Milan, Italy

<sup>2</sup>Department of Electrical and Information Engineering, University of Cassino and Southern Latium, Cassino, Italy

<sup>3</sup>Unit of Computer Systems and Bioinformatics, Department of Engineering, University Campus Bio-Medico of Rome, Italy

<sup>4</sup>City University of Hong Kong

<sup>5</sup>Unit of Advanced Robotics and Human-Centered Technologies, Department of Engineering, University Campus Bio-Medico of Rome, Italy

<sup>6</sup>Shenzhen University

<sup>7</sup>Department of Diagnostics and Intervention, Radiation Physics, Biomedical Engineering, Umeå University, Sweden

## Abstract

In the last few years Artificial Intelligence (AI) is emerging as a game changer in many areas of society and, in particular, its integration in medicine heralds a transformative approach towards personalized healthcare and well-being, promising significant improvements in diagnostic precision, therapeutic outcomes, and patient care. Our research explores the cutting-edge realms of multimodal AI, resilient AI, and healthcare robotics, aiming to harness the synergy of diverse data modalities and advanced computational models to redefine healthcare paradigms. This multidisciplinary effort seeks to bridge technology and clinical practice, advancing AI-driven next generation personalized healthcare and well-being.

## Keywords

Artificial Intelligence, Multimodal Learning, Precision Medicine, Stress Detection, Resilient AI, Healthcare Robotics

## 1. Introduction

Artificial Intelligence (AI) has proven itself as enabling factor for triggering great transformations of society [1, 2, 3, 4]. However on the verge of the fifth industrial revolution, there are several challenges that involve the consolidation of AI arrival in sectors as medicine and people well-being. Indeed this paradigm shift towards AI-driven healthcare is not just a technological revolution; it represents a comprehensive reimagining of medical practices, enhancing the quality, efficiency, and accessibility of healthcare services. In this scenario our efforts are directed towards four research paths: (i) multimodal AI for precision medicine (section 2); (ii) multimodal AI to foster wellbeing (section 3); (iii) resilient AI (section 4); (iv) AI in robotics for healthcare (section 5). For each of these routes we provide a brief description of the developed solutions, highlighting solved problems and open challenges.

## 2. Multimodal AI enables precision medicine

The evolution of precision medicine marks a paradigm shift from the traditional "one-size-fits-all" approach in healthcare towards tailored therapeutic strategies that account for individual variability in genes, environment, and lifestyle. In this context, leveraging the variety of patient generated data (as images, clinical data, electronic health records etc.) can provide a significant boost to unlocking a holistic view of the patient. Towards this end multimodal AI provides the ultimate tool [5, 6]: the integration is not merely additive, it's transformative, enabling the extraction of insights that would remain obscured under traditional, unimodal analysis. We are currently studying the potential of multimodal AI for precision medicine facing different challenges in different application domains: in the oncological domain we face challenges regarding data fusion and representation, with two projects on Non-Small Cell Lung Cancer (NSCLC) (sections 2.1 and 2.2); in augmenting the diagnosis and prognosis of Alzheimer (section 2.3) we tackle the problem of imbalance in multimodal datasets; and in COVID-19 prognosis (section 2.4) we attempt to solve issues related to scarcity of large, labelled datasets

*Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29-30, 2024, Naples, Italy*

\*Corresponding author.

✉ e.cordelli@unicampus.it (E. Cordelli)

🆔 0000-0001-6062-7575 (E. Cordelli)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

with compatible tasks for training deep learning models without leading to overfitting.

## 2.1. AIDA

AIDA stands for "explAinable multiModal Deep learning for personAlized oncology". This project faces the challenge of advancing Multimodal Deep Learning (MDL), studying how to learn shared representations between different modalities, by investigating when to fuse the different modalities and how to embed in the training any process able to learn more powerful data representations. All this is directed towards facing the association between radiomic, pathomic and Electronic Health Records (EHRs) in precision oncology to predict the patient outcomes in terms of progression free survival, overall survival, relapse time and response rate NSCLC, which represents the 85% of all lung cancer cases. To pursue these objectives we started from learning unimodal representation of EHRs and medical imaging.

As a prior contribution, EHRs are vital resources for documenting patient clinical history and procedures, but are often challenging to process due to their unstructured nature. Natural Language Processing (NLP) tools, particularly Named Entity Recognition (NER) with the use of Transformer-based models, have proven effective in extracting meaningful information from EHRs [7]. Transformers excel at capturing contextual relationships between words and the still not thoroughly explored contextual embedding they create can enhance the understanding of the content itself. We propose the Hierarchical Embedding Attention for overall survival (HEAL), a methodology that leverages multi-class NER-driven representations from EHRs by weighting them with attentional mechanisms. The ability of emphasizing clinically relevant information within unstructured data, operating both at word and sentence levels, makes HEAL more interpretable for medical applications. In a NSCLC Overall Survival (OS) prediction case study, HEAL achieved an average  $C^{td}$ -index of 0.639 and a low standard deviation of 0.014 over 5 runs, showing a statistically significant superiority with respect to manually extracted clinical features.

Our second contribution, even if still at its preliminary steps, grounds on the fact that deep learning (DL) approaches have demonstrated significant value in automatically learning potentially relevant patterns from medical images, such as computed tomography (CT) [8]. Hence, in this study we explore a novel methodology for predicting OS in NSCLC patients using only CT images, aiming at a multitask architecture that encompasses prognostic factors like Progression-Free Survival (PFS) beyond predicting OS alone. The first steps in this direction include producing a soft attention weighted feature map for each input slice and highlighting the relevant

slices crucial for predicting OS outcome.

## 2.2. PICTURE

PICTURE stands for "Pathological response AI-driven prediction after neoadjuvant therapies in NSCLC". This project is based on the central hypothesis that heterogeneous medical data (i.e. radiological images, histology images, cytology and molecular data and EHRs) are consistent with the pathological complete response (pCR), so their combination using artificial intelligence (AI) can provide accurate pCR prediction in NSCLC patients. Indeed, albeit treating locally advanced NSCLC surgically is the mainstay, it is important to prevent post-surgery recurrence, and neoadjuvant therapy (NAT) has shown potential in enhancing overall survival rates and achieving a complete pathological response, that, if correctly evaluated before the treatment, can even avoid non necessary surgical resections.

PICTURE pursues three objectives: (i) pCR prediction through radiology imaging, histology, cytology, molecular data, EHRs, and their combination; (ii) leveraging multimodal deep learning to make the performance of AI resilient and robust for pCR prediction signature; (iii) improving trust and transparency using explainable AI models. PICTURE also has the exploratory aim of transferring trained models to predict pCR for patients undergoing chemoimmunotherapy, tailoring treatments to the individual needs of patients.

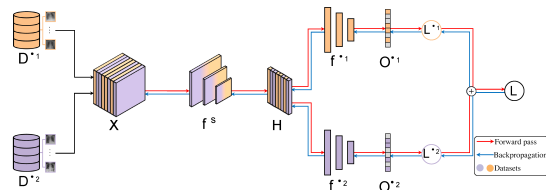
## 2.3. Facing imbalance in Alzheimer's Disease diagnosis and prognosis

Alzheimer's disease (AD) is a progressive neurodegenerative condition with decline in cognitive function, and because of the lack of a cure, its early detection is paramount. Despite the recent progress in AI, challenges such as class imbalance, integration of multimodal data, and robust generalization remain pervasive. In response to this we introduce a novel methodology that leverages the strengths of ensemble learning while incorporating advanced fusion techniques. For each of the 4 modalities of the tabular ADNI database, we train a series of classifiers on varied class distributions followed by a late fusion strategy that integrates the different modalities to improve the results.

Our framework is evaluated on two diagnostic tasks (binary and ternary) and four binary prognostic tasks (at 12, 24, 36, and 48 months) and compared with 12 state-of-the-art imbalanced data algorithms, achieving 97.04% g-mean on the binary diagnostic task and 90.81% g-mean on the 48-month prognostic task.

## 2.4. Multi-Dataset Multi-Task Learning for COVID-19 Prognosis

In COVID-19 context [9] in order to fight the scarcity of large, labelled chest radiographic images (CXR) datasets, we introduce a novel multi-dataset multi-task (MDMT) training framework, by integrating correlated datasets from disparate sources and assessing severity score to classify prognostic severity groups [10, 11, 12], instead of relying on datasets with multiple and correlated labelling schemes. As illustrated in figure 1 a deep CNN takes the images as input and branches into task-specific fully connected output networks, to end with a multi-task loss function incorporating an indicator function to exploit multi-dataset integration.



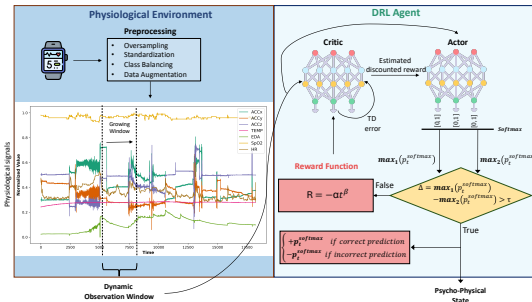
**Figure 1:** Overview of the proposed Multi-Dataset Multi-Task model architecture, composed by a shared backbone  $f^s$  and two task-specific fully connected network heads,  $f^{T1}$  and  $f^{T2}$ , for tasks  $\tau_1$  and  $\tau_2$ , respectively, producing outputs  $O^{T1}$  and  $O^{T2}$ .

Proceeding with a 5 cross-validation and leave one center out training, we evaluated the method across 18 different CNN backbone on prognosis classification task and fine-tuning from BRIXIA dataset to AIforCOVID dataset task. Best average performance with statistical robustness achieved: 68.6% accuracy, 66.6% F1-score and 68.5% g-mean for the 5 cross validation, and 65.7% accuracy, 64.3% F1-score and 66.0% g-mean for the leave-one-center-out validation strategy. Future directions include new domains and the integration of XAI [6].

## 3. Multimodal AI to foster well-being

Stress, a response to physical and emotional demands, is crucial in determining individuals' well-being and if unmanaged can lead to conditions such as anxiety, depression and cardiovascular diseases. Also in this scenario multimodal AI offers a tool for proactive approach to health management, in order to provide real-time monitoring and interventions, thereby mitigating long-term health risks associated with chronic stress.

We are targeting stress detection from two perspectives: first, we are focusing on maximising the stress level prediction accuracy within the shortest possible time, exploiting multimodal physiological time series data and



**Figure 2:** Overview of the proposed method for early stress detection consisting of two main blocks: physiological environment and DRL agent. The first involves the pre-processing of data and the description of the dynamic observation space. The second block incorporates a SAC-based DRL agent fed with data from the first block to control the system.

Deep Reinforcement Learning (DRL); second, we are further expanding the multimodal view integrating information from video, audio and text with the physiological data. Robust and fast stress detection approaches can bring benefit in several contexts: from providing a targeted and more personal assistance to patients, to ensuring safety for workers, for instance, Air Traffic Controllers (ATC) that endure high levels of psychological pressure during their job impacting operational safety.

Our first approach [13] employs a new DRL model to identify stress indicators. We obtained this by leveraging a dynamic time observation window that expands each step of the learning process, asking the agent to choose either to continue observing or to classify based on the information gathered until that point, trying to minimize the amount of data required for decision-making. As depicted in the figure Figure 2, we adopted the Soft Actor-Critic algorithm for its effectiveness in handling continuous action spaces. In a Leave-One-Subject-Out approach with data augmentation on the Non-EEG public dataset we outperformed existing solutions, showing the power of DRL for early stress detection.

On top of this approach we are exploring the larger multimodal asset of the Ulm-Trier Social Stress Test dataset (ULM-TSST, MuSe 2022 challenge), containing 41 training, 14 validation and 14 test subjects, simulating a job interview scenario, with audio, video, text, and physiological data modalities, rated on arousal and valence stress parameters. The aim is to build a high performance architecture that leverages non invasive modalities for stress detection that can be employed in work environments. In both cases the scarcity of large datasets that provide a quantitative measurement of stress is still the main challenge: we will try to face it considering the construction of robust and specific acquisition protocols to test the effectiveness of the developed approaches in

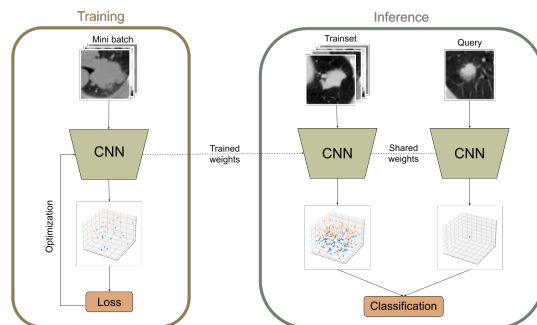
real-world scenarios.

## 4. Resilient AI

Due to the high stakes involved in healthcare decisions, the sensitivity of medical data, and the complexity of medical environments, AI systems should be designed to maintain their intended performance and integrity in the face of adversities, as data corruption, missing data, privacy leakage, or unexpected changes in their operating environment. This is the goal of Resilient AI, which is an aspect that cannot be left out when the aim is to integrate AI for augmenting the medical practice. To meet this goal we are currently investigating three main aspects that fall under the Resilient AI umbrella: developing systems robust to missing data, the challenge of limited extension datasets and how to protect sensitive patients data.

With respect to the missing data challenge [14], although a variety of strategies exist for addressing this problem in health datasets, to overcome the obstacle to select the most suitable one and their dependency on the dataset's specifics, we developed a Transformer-based model [15] that applies masking to ignore the missing data, thus eliminating the need of imputation and deletion techniques and focusing directly only on the available features through self-attention. Moreover, we introduced a novel feature-identifying form of positional encoding to facilitate the integration of tabular data into a Transformer framework. This method was validated through an overall survival classification task, employing clinical data from the CLARO [16] project and improving the prediction accuracy.

In order to address the problem of working with datasets limited in the extension, particularly frequent in healthcare domain, Triplet networks, a subtype of the Siamese networks, emerge as a promising solution, comprising three identical networks operating concurrently. Throughout training of these three networks two inputs belong to the same class, whereas the third belongs to a distinct class, with the final objective to develop a feature space with two distinct clusters one per class by incorporating inter-class diversities alongside intra-class similarities and providing scenarios with limited data with more triplets compared to instances (Figure 3). In our study [17], using a private dataset of 86 CT scans, triplet networks surpass the plain deep networks in accurately predicting the histological subtypes of NSCLC patients. Currently, we are broadening the scope of our research including PET images alongside CT scans and adopting a multimodal strategy for the same classification. By integrating these complementary data we anticipate achieving a significant improvement and overcoming the challenges posed by limited data scenarios.



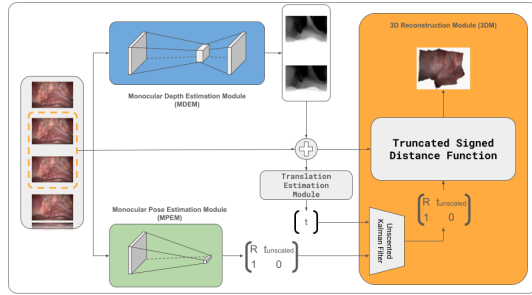
**Figure 3:** Overall framework of the proposed method working with triplet networks.

Last but not least, the challenge related to patient privacy led us to explore Federated learning (FL). FL presents an innovative solution to the challenge of protecting sensitive patient data in artificial intelligence applications in healthcare, in fact enabling the training of a shared global model with a central server while ensuring data privacy within local institutions. On this basis we introduce a new token-based FL paradigm, revolutionizing the traditional approach with sequential or random passing of a token between clients during each epoch. This innovative method allows only the token owner to send the weights to the server, which redistributes them directly to all models. By eliminating local training epochs and allowing immediate transmission, this paradigm shift streamlines the process by circulating a single model among clients and also mitigating the need for an initial warm-up period, potentially paving the way for a decentralized system that reduces dependence on a central server and minimizes the number of parameters transmitted in each iteration. Results on the tabular part of the AIforCOVID dataset [18] composed of 6 hospitals show that the performance of the FL model does not deviate from that of its equivalent trained on all datasets aggregated into a single pool. The next steps will focus on integrating other modalities into the FL pipeline, such as CXR scans of the AIforCOVID dataset itself.

## 5. AI for healthcare robotics

The integration of robotics in healthcare settings exemplifies another dimension of AI's impact, automating routine tasks, assisting in surgeries with precision beyond human capability, and providing rehabilitation support to patients. This not only enhances service delivery but also alleviates the workload on healthcare professionals, allowing them to focus more on patient-centered care. In this scenario we pursue two aims: first, enhancing robotic surgery with real-time high precision localiza-





**Figure 4:** The MVSLAM pipeline integrates depth estimation, pose estimation, and 3D reconstruction modules to generate a continuously updated 3D map of the surgical environment from monocular endoscopic video frames.

tion; second, boosting lower-limb robotic rehabilitation optimizing the structural exoskeleton sensor configuration.

For the first objective we focus on the laparoscopy use case, as one of the preferred surgical methods. Despite recent advancements in image acquisition it is still limited to rely on 2D images view: misinterpreting anatomical structures due to this limit is a common source of errors. In contrast, 3D imaging increases the accuracy of instrument manipulation, leads to better outcomes in surgery, and shortens the learning for trainees. Even several research in surgical 3D imaging has been explored, like camera-based tracking and mapping, Mosaicking, Structure from Motion, and Shape from Template, they often rely on simplifications that can limit their effectiveness. On this ground Simultaneous Localization and Mapping (SLAM) has shown promising results, as it aims to create a map of the environment while localizing the sensor position within it. Therefore we developed a robust deep learning SLAM pipeline to operate in real-time across diverse surgical settings by providing an immersive, interactive 3D environment (Figure 4), allowing for more precise and personalized interventions with the future possibility to be integrated with augmented reality displays.

For the second objective, we focus on the challenges in the field of lower limb robotics. It aims at supporting people with lower limb disabilities by enhancing movement, mobility, and providing targeted exercise. Technologies as exoskeletons, prosthetics, and rehabilitation robots, are particularly helpful for those with neurological issues, offering improved rehabilitation, independence, and tailored care. Effective use requires precise control settings to adapt walking patterns to different terrains. Challenges in this field involve the extensive need for sensors for terrain detection and the complexity in processing sensor data. Simplifying sensor requirements to accurately determine terrain and slope is critical for user-

friendly, efficient operation, and safety. The aim of our work is to recognize the terrain on which an exoskeleton is walking and its inclination. Among several state-of-the-art driven approaches, we achieved promising results using LSTM architectures with IMU data (0.94 of accuracy in Leave-one-out cross-validation), and CNN-LSTM architectures with EMG data (0.75 of accuracy). The fusion of IMU and EMG data didn't bring any significant improvement, as explanatory tests indicated that the best 20 contributing features belong to IMU. Next, by varying the number of sensors, and therefore features, we noticed that the best results are achieved by selecting the most relevant features, from one to three, according to SHAP (on a 3 subjects validation set), leading to 0.85, 0.89 and 0.93 of accuracy respectively. Lastly, we found that LSTM and CNN-LSTM are valid architectures for slope inclination prediction (MAE of  $1.95^\circ$ ) and stair height (MAE of 15.65 mm), without significant differences in employing 3 or 4 sensors.

## Acknowledgments

Fatih Aksu, Alice Natalina Caragliano, Camillo Maria Caruso, Omar Coser, Arianna Francesconi, Leonardo Furlia, Guido Manni, Giustino Marino, Domenico Paolo and Filippo Ruffini are Ph.D. students enrolled in the National Ph.D. in Artificial Intelligence, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma. We acknowledge financial support from: i) PNRR MUR project PE0000013-FAIR; ii) PRIN 2022 MUR 20228MZFAA-AIDA (CUP C53D23003620008); iii) PRIN PNRR 2022 MUR P2022P3CXJ-PICTURE (CUP C53D23009280001); iv) FCS MISE (CUP B89J23000580005); v) MAECI (grant n. CN23GR09); vi) NRR MUR project PNC0000007 Fit4MedRob. This work was also partially supported by the following companies: Eustema S.p.A. and ENAV S.p.A..

## References

- [1] V. Guarrasi, L. Tronchin, C. M. Caruso, A. Rofena, G. Manni, F. Aksu, D. Paolo, G. Iannello, R. Sicilia, E. Cordelli, et al., Building an ai-enabled metaverse for intelligent healthcare: opportunities and challenges, in: *Ital-IA 2023, Italia Intelligenza Artificiale Thematic Workshops*, co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023, CEUR-WS, 2023, pp. 134–139.
- [2] E. Cordelli, V. Guarrasi, G. Iannello, F. Ruffini, R. Sicilia, P. Soda, L. Tronchin, Making ai trustworthy in multimodal and healthcare scenarios, *Proceedings of the Ital-IA (2023)*.

- [3] J. P. Bharadiya, Artificial intelligence and the future of web 3.0: Opportunities and challenges ahead, *American Journal of Computer Science and Technology* 6 (2023) 91–96.
- [4] V. Pereira, E. Hadjielias, M. Christofi, D. Vrontis, A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective, *Human Resource Management Review* 33 (2023) 100857.
- [5] V. Guarrasi, P. Soda, Multi-objective optimization determines when, which and how to fuse deep networks: An application to predict covid-19 outcomes, *Computers in Biology and Medicine* 154 (2023) 106625.
- [6] V. Guarrasi, L. Tronchin, D. Albano, E. Faiella, D. Fazzini, D. Santucci, P. Soda, Multimodal explainability via latent shift applied to covid-19 stratification, *arXiv preprint arXiv:2212.14084* (2022).
- [7] D. Paolo, et al., Named entity recognition in italian lung cancer clinical reports using transformers, in: *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2023, pp. 4101–4107.
- [8] M. Tortora, et al., Radiopathomics: multimodal learning in non-small cell lung cancer for adaptive radiotherapy, *IEEE Access* (2023).
- [9] G. Fiscon, F. Salvatore, V. Guarrasi, A. R. Garbuglia, P. Paci, Assessing the impact of data-driven limitations on tracing and forecasting the outbreak dynamics of covid-19, *Computers in biology and medicine* 135 (2021) 104657.
- [10] V. Guarrasi, N. C. D’Amico, R. Sicilia, E. Cordelli, P. Soda, A multi-expert system to detect covid-19 cases in x-ray images, in: *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2021, pp. 395–400.
- [11] V. Guarrasi, N. C. D’Amico, R. Sicilia, E. Cordelli, P. Soda, Pareto optimization of deep networks for covid-19 diagnosis from chest x-rays, *Pattern Recognition* 121 (2022) 108242.
- [12] V. Guarrasi, P. Soda, Optimized fusion of cnns to diagnose pulmonary diseases on chest x-rays, in: *International Conference on Image Analysis and Processing*, Springer, 2022, pp. 197–209.
- [13] L. Furia, et al., Exploring early stress detection from multimodal time series with deep reinforcement learning, in: *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2023, pp. 1917–1920.
- [14] A. Rofena, V. Guarrasi, M. Sarli, C. L. Piccolo, M. Sammarra, B. B. Zobel, P. Soda, A deep learning approach for virtual contrast enhancement in contrast enhanced spectral mammography, *arXiv preprint arXiv:2308.00471* (2023).
- [15] C. M. Caruso, V. Guarrasi, S. Ramella, P. Soda, A deep learning approach for overall survival analysis with missing values, *arXiv preprint arXiv:2307.11465* (2023).
- [16] CLARO - CoLLaborative multi-sources Radiopathomics approach for personalized Oncology in non-small cell lung cancer., <http://www.cosbi-lab.it/claro/>, 2020. Accessed: 2023-03-20.
- [17] F. Aksu, et al., Early experiences on using triplet networks for histological subtype classification in non-small cell lung cancer, in: *2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2023, pp. 832–837.
- [18] P. Soda, N. C. D’Amico, J. Tessadori, G. Valbusa, V. Guarrasi, C. Bortolotto, M. U. Akbar, R. Sicilia, E. Cordelli, D. Fazzini, et al., Aiforcovid: Predicting the clinical outcomes in patients with covid-19 applying ai to chest-x-rays. an italian multicentre study, *Medical image analysis* 74 (2021) 102216.