Artificial intelligence methods for biomedical imaging and omics data

Carlo Alberto Barbano[1,2], Marco Beccuti[1], Francesca Cordero[1], Desislav Nikolaev Ivanov[1], Nicola Licheri[1], Simone Pernice[1], Alberto Presta[1], Riccardo Renzulli[1] and Marco Grangetto[1,*]

[1]Computer Science dept., University of Turin, Italy
[2]LTCI, Télécom Paris, IP Paris, France

Abstract

The use of deep learning in biomedical imaging and omics data has shown great potential for enhancing medical diagnosis and improving patient outcomes. In this paper, we present the deep learning and machine learning research activities of two research laboratories: EIDOS and qBio. Our research encompasses a broad range of topics, including digital pathology, integration of omics data, digital radiology, computational epidemiology and neuroimaging. We collaborate with several hospitals for the collection of relevant datasets and with international research centers and foreign universities to develop state-of-the-art techniques. Overall, we believe that the activities of these laboratories in deep and machine learning have the potential to improve the way we diagnose and treat various medical conditions.

Keywords

Biomedical imaging, deep learning, integration of omics data, histopathology, radiology, neuroimaging, machine learning, computational epidemiology

1. Introduction

In this paper we describe some recent research lines in the wide area of biomedical image processing and machine learning approaches to analyze omics data explored by the EIDOS lab [1] and qBio lab [2] at the Computer Science Department of the University of Turin, respectively. EIDOS lab is also a member of the Italian Association for Computer Vision, Pattern Recognition and Machine Learning [3], and qbio lab is involved in the steering committee of the Bioinformatics Italian Society (BITS) [4] and of CINI InfoLife laboratory [5].

The experience of EIDOS in biomedical image processing and analysis is grounded in several projects and collaboration with major hospitals. EIDOS was recently supported by the EU through the DeepHealth project [6] *Deep-Learning and HPC to Boost Biomedical Applications for Health* and currently by Regione Piemonte through the Co.R.S.A. [7] (Covid Radiographic imaging System based on AI). Within these projects, EIDOS gained experience in exploiting Deep Learning in several clinical use cases with the publication of 4 open datasets in different medical domains, namely histopathology, radiology and neurology. The Co.R.S.A. project is currently moving into the evaluation of the clinical impact of AI tools to support radiological diagnosis of COVID-19.

The expertise of the qBio group covers different areas of bioinformatics as shown by the various research projects in which the qBio's members are actually involved. For instance, in the ONCOBIOME project [8] *Gut OncoMicrobiome Signatures (GOMS) associated with cancer incidence, prognosis and prediction of treatment response* funded by the EC under the topic SC1-BHC-03-2018, the qBio group provided the analysis and integration of omics data (metagenome, small noncoding RNAomics, transcriptomic, metabolomics) by machine learning techniques to identify diagnostic and prognostic biomarker in four solid cancers. In the *Minimal residual disease in follicular and mantle cell lymphoma: development and validation of novel tools and predictive models* project funded by the Italian Ministry of Health (Finalizzata 2021) the qBio group developed an innovative approach for clustering functional data to post-treatment outcome predictor in blood cancers. Differently in the SUS-MIRRI.IT project funded by the Italian government on the NextGeneration EU-funded Recovery and Resilience National Plan (PNRR) – Research Infrastructure - to support the Italian network of collections of microorganisms the qBio group works on the implementation of the computational infrastructure enabling the IT activi-

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
carlo.barbano@unito.it (C. A. Barbano); marco.beccuti@unito.it (M. Beccuti); francesca.cordero@unito.it (F. Cordero); desislav.ivanov@edu.unito.it (D. N. Ivanov); nicola.licheri@unito.it (N. Licheri); simone.pernice@unito.it (S. Pernice); alberto.presta@unito.it (A. Presta); riccardo.renzulli@unito.it (R. Renzulli); marco.grangetto@unito.it (M. Grangetto)
0000-0001-9512-0440 (C. A. Barbano); 0000-0001-6125-9460 (M. Beccuti); 0000-0002-3143-3330 (F. Cordero); 0009-0007-4624-2506 (D. N. Ivanov); 0000-0003-1075-7333 (N. Licheri); 0000-0001-7124-4676 (S. Pernice); 0000-0001-6590-3604 (A. Presta); 0000-0003-0532-5966 (R. Renzulli); 0000-0002-2709-7864 (M. Grangetto)

ties of the project. Finally, in TrustAlert project, founded by "La Compagnia di San Paolo" and "Fondazione CDP" (grant dedicated to AI) the qBio group works on developing a smart platform for providing early warnings, monitoring, and forecasting tools for public health response agencies and local healthcare services for anticipating medical needs.

2. Digital pathology

Our group has been actively engaged in exploring the potential of deep learning techniques for digital pathology applications. In this section, we present an overview of our group activities in this domain, focusing on three key areas of research. Firstly, we discuss our efforts in creating a large dataset of whole-slide images of colorectal polyps for diagnostic purposes. Secondly, we delve into our work on generative models for synthetic data augmentation to enhance the accuracy and robustness of our models. Finally, we describe our efforts in developing deep learning-based methods for grading colorectal cancer. Our goal is to provide a comprehensive overview of our group's contributions in this field and to highlight the potential of deep learning for advancing digital pathology.

2.1. The UnitoPatho data collection

Digital histopathology solutions have gained increasing demand, fuelued by the widespread adoption of cancer screening programs [9]. In particular we have been challenged by gastrointestinal histopathologists, who inspect tissue samples collected during colonoscopies, to provide automatic tools to recognize and classify colorectal polyps. Colorectal polyps are pre-malignant lesions that are analyzed to *i)* classify the polyp type (hyperplastic, adenoma) and *ii)* to evaluate the dysplasia grade in case of adenomas. In this field the search for computer aided solutions is of paramount importance not only for the common need to simplify and speed-up the pathologists' clinical routine. Indeed, the concordance rate among pathologists is difficult to guarantee: for instance, the concordance in assessing a tubulo-villous polyp or low grade dysplasia is reported to be around 70% [10]. From the technical point of view, the nature of this task poses a number of challenges that must be taken into account, which have been described in our recent publications [11, 12]. The first issue to overcome, as usual, is the scarce availability of data. We tackled this issue by building and releasing *UniToPatho* [12, 13], a high-resolution annotated dataset of Hematoxylin and Eosin (H&E)-stained colorectal patches extracted from whole-slide images (WSI). Another challenge is posed by the huge resolution of WSIs, which makes the application of standard DL classification pipelines not straightforward due to computational requirements. Also, when dealing with H&E-stained images, it is often necessary to account for the different concentrations of the two histological stains, depending on how the staining procedure was carried on. This is why we also developed *torchstain* a popular stain normalization tool aimed at DL applications [14] which has increasingly gained traction in the community.

2.2. Generative models for synthetic data augmentation

Accurately grading dysplasia presents a major challenge due to the underrepresented high-grade class in the UniToPatho dataset. To address this, we explored the use of Generative Adversarial Networks (GANs) to generate new samples for augmentation. In our initial experiment, we trained StyleGANs on different resolutions using both unconditional (one GAN for each class) and conditional (by conditioning on the grade) settings. Our analysis revealed intriguing details of the Generator network, such as high-quality generations almost indistinguishable from real ones, as well as latent-space properties that determine the distribution and positioning of cells in the output images. Surprisingly, we discovered that, with the same method described in [12], training a ResNet for classification solely on synthetic data produced nearly the same results as the real-data approach, with just 1.5% lower accuracy on the real test set. However, despite augmenting the real dataset with multiple variants of synthetic samples, we did not significantly improve the classification accuracy of the ResNets. The strong imbalance in the dataset hindered our GANs from learning the real distribution of the high-grade class. To overcome this, we developed a novel StyleGAN architecture that guides the synthesis process using high-level tissue features defined through segmentation masks of nuclei. We created a new GAN based on the initial StyleGAN that uses an additional UNet-like architecture for injecting the segmentation masks into the synthesis network. Our new model achieved competitive FID, comparable with our initial GAN approach. Additionally, we built a tool that allows editing segmentation masks and visualizing the generated results in real time. We plan to use these novel models and tools for a more precise augmentation of the underrepresented class, by exploiting expert medical knowledge when manipulating the nuclei masks.

2.3. Colorectal cancer grading integrating heterogeneous features

At the basis of the personalized medicine approaches for the prediction of stage and for the prevention of the disease there are new AI computational approaches based on

the exploitation and integration of data from omics studies. The identification of patients with a poor prognosis, nonresponders to standard therapies, or with an elevated probability to have an adverse effect is the main goal of the integrative omics data approach. The identification of a heterogeneous signature based on transcriptomics, metabolomics, metagenomics data, and the imaging features extracted from cancer slides could be used as predictive of the progression of cancer and therefore adapt preventive therapies. The signature identification is generally obtained by the exploration of both early and late integration strategy [15]. The derived signatures will be built either on a combination of individual features (such as expression level or abundance/presence of a metabolite) or of composite features - summary values of groups of highly correlated variables.

3. Digital Radiology

In recent years, digital radiology has seen a surge in research and development due to the advancements in machine learning and computer vision techniques. Our group has been actively involved in exploring various aspects of digital radiology, ranging from Covid-19 detection to lung nodule detection and calcium score prediction. In this section, we will discuss our group's activities and contributions in these areas, highlighting the C.o.R.S.A. Project for Covid-19 detection, the UnitoChest dataset for lung nodule detection, and our efforts in calcium score prediction from CXR. These projects represents a significant step forward in the field of digital radiology and have the potential to improve patient outcomes by enabling earlier and more accurate diagnoses.

3.1. COVID-19 and the C.o.R.S.A. project

At the peak of the Covid-19 pandemic in Italy in 2020, our group started a collaboration with the radiology units of local hospitals to help with the screening of Covid-19 patients. The first results of our effort toward Covid-19 detection from chest X-ray (CXR) can be found in [16, 17]. Beside the crude achievements in terms of detection performance (sensitivity and specificity consistently above 0.7) these efforts have contributed to highlight other fundamental aspects, namely the difficulty to cope with small and imbalanced datasets. In particular, our Covid-19 study has exacerbated the difficulty to cope with small data since the data collection was in progress at emergency time during our studies, an issue that affected most of the studies at that time. In fact, during the pandemic, it was quite impossible to get balanced and unbiased samples, e.g. the large majority of the admitted patients were actually affected by Covid-19, and using publicly available data presented many challenges as described in [16, 17]. Our collaboration with the radiology units of Città della Salute e Della Scienza (CDSS), Mauriziano, San Luigi, Monzino and ASLTo3 have evolved into the regionally funded C.o.R.S.A. project[1]. The joint effort on this topic led to the CORDA data collection, which is publicly available for download[2], and contains around 3000 images from patient who underwent Covid-19 screening, along with the ground-truth label obtained with RT-PCR testing (swab). Our ongoing efforts focus on obtaining models which are robust to biases in the data [18], robust to noise given by different acquisition sites and which can provide some form of explainability. For the latter, in [17] a DL diagnostic approach that imitates the radiologist diagnosis process is proposed, based on a preliminary classification stage mapping onto standard radiological findings from the lungs, on top of which the Covid-19 is diagnosed.

3.2. Lung nodules segmentation and the UnitoChest dataset

Lung cancer is the primary cause of death for men and women, with a survival rate lower than breast and prostate cancer [19]. Therefore, early detection of lung nodules is the key to early cancer diagnosis and treatment effectiveness assessment. Deep neural networks achieve outstanding lung nodules detection, classification, and segmentation results. However, the quality and quantity of the training images can boost their performance. Within the DeepHealth project, we created UniToChest [20], a dataset consisting Computed Tomography (CT) scans of 623 patients. UniToChest is publicly available[3] and is the largest of its kind and boasts a diversity of patient ages, acquisition machines and nodules diameters. Manual lung nodules segmentation is time-consuming and prone to errors; so, several systems based on deep learning have been proposed for the detection and segmentation of lung nodules. In our study [21], we analyzed a U-Net based architecture that yields promising results in both detection and segmentation of lung nodules. Future research directions of this work include exploiting the three-dimensional information of nodules across neighboring slices.

3.3. Calcium score prediction

Coronary artery disease is the leading cause of death in industrialized countries, despite significant advances in diagnosis and therapy. In particular, It is now known that coronary calcium, indicated with a value called *calcium*

[1] https://corsa.di.unito.it/. Project funded by Regione Piemonte - Bando INFRA-P2.
[2] https://zenodo.org/record/7501816
[3] https://zenodo.org/record/5797912

score (CAC), is associated with sub-clinical atherosclerotic diseases, since its absence is associated with an extremely low and long-lasting probability of cardiovascular events even in subjects at high risk [22]. Calcifications of the coronary arteries are easily detectable through CT scans, while it is difficult to visually detect them from an X-ray image, even for an expert radiologist: It is precisely for this reason that an automatic system for analyzing X-ray images could be of fundamental importance for the early detection of calcium. Up to our knowledge, there is only one study where the presence of calcium is inferred from X-rays [23]. In our project we studied the possibility to train properly a CNN able to detect the presence of calcium in coronary arteries using Chest X-rays as input data instead of CT scan. In particular, Our model first tries to give an accurate prediction of the value of coronary artery calcium (CAC), and based on the latter, it determines whether or not a patient has coronary calcium. This project was developed in collaboration with the radiology unit of Città della Salute e della Scienza di Torino (CDSS) hospital in Turin, Who collected a dataset of 506 chest X-rays, equally divided between patients with and without coronary calcium.

4. Neuroimaging

Our research group is also focused on developing innovative methods and tools for neuroimaging analysis, with a particular emphasis on Deep Learning (DL) approaches. In this paper section, we present two subsections showcasing our recent work in the field of neuroimaging. The first subsection focuses on brain age prediction from MRI, a challenging task that requires robust and accurate models capable of generalizing across different imaging sites. The second subsection explores the use of DL techniques for the generation of brain perfusion maps from CT images, aiming to improve the diagnosis and treatment of ischemic stroke.

4.1. Brain age prediction

Brain aging involves complex biological processes, such as cortical thinning, that are highly heterogeneous across individuals, suggesting that people do not age in the same manner. Accurately modeling brain aging at the subject-level is a long-standing goal in neuroscience as it could enhance our understanding of age-related diseases such as neurodegenerative disorders. To this end, brain-age predictors linking neuroanatomy to chronological age have been proposed using Deep Learning (DL) [24]. In order to build accurate biomarker of aging, DL models need large-scale neuroimaging dataset for training, which often involves multi-site studies, partly because of the high cost per patient in each study. Recent works have shown that DL models, and in particular Deep Neural Networks (DNN), largely over-fit site-related noise when trained on such multi-site datasets, notably due to the difference in acquisition protocols, scanner constructors, physical properties such as permanent magnetic field [25, 26]. This also implies poor generalization performance on data from new incoming sites, highly limiting the applicability of these models to real-life scenarios. In order to build more accurate brain age models, the OpenBHB challenge [27] has been recently released[3]. It is an open-ended challenge, publicly available, which provides one of the largest datasets of healthy brain MRIs. In this context, together with our partners at Télécom Paris and NeuroSpin, CEA, we have developed a novel contrastive learning loss for regression of brain age from MRI [28]. We validated it on the OpenBHB challenge, where chronological age must be learned without being affected by site-related noise. With our method, we obtain the best results on the official challenge leaderboard.

4.2. Brain perfusion and UnitoBrain

CT brain imaging and in particular CT perfusion (CTP) has become established tool in treatment of ischaemic stroke. During CTP, a series of low-dose scans are acquired after contrast bolus injection, allowing to compute parametric maps to track perfusion parameters dynamics, e.g. Cerebral Blood Volume (CBV). In our study [29] we explored whether a properly trained CNN, based on a U-Net-like structure, can generate informative, parametric maps such as CBV. The UNITOBrain dataset [30] we created to support the research is publicly available[4] and attracted considerable interest in the area. In the end, the agreement between our CNN-based perfusion maps and the state-of-the-art perfusion analysis methods maps based on deconvolution of the data highlights the potential of deep learning methods applied to perfusion analysis. Moreover, machine learning methods can reduce data inputs required to estimate the ischemic core and thus might allow the development of novel perfusion protocols with lower radiation dose.

5. Omics data for biomarkers discovery and computational epidemiology

5.1. Dataset

In collaboration with the Italian Istitute for Genomic Medicine (IIGM) [31] we collected samples (i.e. stool, cancer tissue and adjacent tissues, plasma) from patients

[3] https://baobablab.github.io/bhb/
[4] https://ieee-dataport.org/open-access/unitobrain

at the Clinica S. Rita in Vercelli, Italy. Patients with hereditary CRC syndromes, with a previous history of CRC, and with uncompleted or poorly cleaned colonoscopy, were excluded from the study. Patients were recruited at initial diagnosis and had not received any treatment prior to fecal sample collection. Subjects reporting the use of antibiotics during the 6 months prior to the sample collection were excluded from the study. On the basis of colonoscopy results, recruited subjects were classified into three categories: 1) healthy subjects: individuals with colonoscopy negative for tumor, adenomas, and other diseases; 2) adenoma patients: individuals with colorectal adenoma/s; and 3) CRC patients: individuals with newly diagnosed CRC. A total of 93 subjects were initially recruited, and the 80 that passed quality control are divided into 29 CRC patients, 27 adenomas, and 24 controls. On the samples collected, the shotgun and small noncoding RNA sequencing were performed.

The main results are reported firstly in [32] where we identified a specific signature composed of profiles of human small non-coding RNAs, microbial sRNAs, and microbial DNAs was able to accurately classify the three categories of subjects with a high level of performance.

This evidence was confirmed across multiple cohorts. Indeed, we assessed the CRC-associated gut microbiome and its ability to distinguish newly diagnosed CRC patients from tumor-free controls. Our study [33] was performed across nine multiple datasets and a combined analysis based on Random Forest based machine learning approach. The identification of reproducible microbial biomarkers for CRC may enable the design of non-invasive diagnostic tools.

5.2. Multi-omics data integration

Deep sequencing technologies allow the production of huge amounts of different omics, each one providing complementary perspectives of the biological system under study. Omics data analysis is challenging because of its high dimensionality, noisiness, and error-proneness. Moreover, most diseases and phenomena affect complex molecular pathways where multiple omics interact with each other, and multi-omics integration can provide a more comprehensive understanding of the biological system under study. Machine learning provides the methodologies to efficiently tackle all these challenges. However, it is difficult to choose the most appropriate algorithm for the data collected, and for the subsequent data integration. The main challenges of such a task are linked to the complexity, heterogeneity, dynamics, uncertainty and high dimensionality, as well as to the right methodologies to analyze and integrate such data.

In this contest, the q-Bio group developed a modular framework for multi-omics integration, called FeatSEE (*Feature Selection, Evaluation, and Explanation*), paper in preparation). User-defined pipelines of analysis can be built by combining the available modules, that implement high-level tasks such as (i) exploratory data analysis, (ii) feature selection for biomarker discovery, (iii) evaluation, for running classification tasks using specific features and learning algorithms, and (iv) feature extraction, for building models based on human-interpretable features in the form of logical rules. The basic idea is to provide a tool for automated machine learning usable by researchers without computational skills, in order to help them in the functional interpretation of results from a clinical/biological point of view.

5.3. Computational Epidemiology

Computational epidemiology exploits Artificial Intelligence and Simulation to successfully support epidemiologists, healthcare professionals, and decision-makers to understand and control the spatio-temporal spread of infectious diseases. In particular, during the first phase of the COVID-19 outbreak, the qBio group was involved, in collaboration with the Department of Medical Sciences of the University of Turin, to support decision-makers of the Italian Piedmont region for evaluating the impact of different implementations of the infection control measures [**?**] (e.g., non-pharmaceutical interventions, surveillance methods, and screening tests) by exploiting the general modelling framework **GreatMod** [34]. Most recently qBio's members are developing an integrated smart dashboard for providing early warnings, monitoring and forecasting tools to public health response agencies and local healthcare services for anticipating medical needs.

References

[1] EidosLab, Image proecessing, computer vision and virtual reality, http://eidos.di.unito.it, 2021.

[2] qBioGroup, Quantitative Biology, http://qbio.di.unito.it, 2022.

[3] CVPL, Italian Association for Computer Vision, Pattern Recognition and Machine Learning, http://www.cvpl.it, 2021.

[4] BITS, Bioinformatics Italian Society, https://bioinformatics.it/, 2000.

[5] InfoLife, CINI national laboratory, https://www.consorzio-cini.it/index.php/it/lab-infolife, 2000.

[6] DeepHealth, Deep-Learning and HPC to Boost Biomedical Applications for Health, https://deephealth-project.eu, 2021.

[7] Co.R.S.A., Covid Radiographic imaging System based on AI, https://corsa.di.unito.it, 2022.

[8] Oncobiome, Gut OncoMicrobiome Signatures (GOMS) associated with cancer incidence, prog-

nosis and prediction of treatment response., https://www.oncobiome.eu/, 2019.
[9] R. S. Gonzalez, Updates and challenges in gastrointestinal pathology, Surgical Pathology Clinics 13 (2020) ix.
[10] B. Denis, C. Peters, C. Chapelain, I. Kleinclaus, A. Fricker, R. Wild, B. Auge, I. Gendre, P. Perrin, D. Chatelain, et al., Diagnostic accuracy of community pathologists in the interpretation of colorectal polyps, European journal of gastroenterology & hepatology 21 (2009) 1153–1160.
[11] D. Perlo, E. Tartaglione, L. Bertero, P. Cassoni, M. Grangetto, Dysplasia grading of colorectal polyps through convolutional neural network analysis of whole slide images, in: Proceedings of Int. Conf. on Medical Imaging and Computer-Aided Diagnosis, 2022, pp. 325–334.
[12] C. A. Barbano, D. Perlo, E. Tartaglione, A. Fiandrotti, L. Bertero, P. Cassoni, M. Grangetto, Unitopatho, a labeled histopathological dataset for colorectal polyps classification and adenoma dysplasia grading, in: IEEE ICIP, 2021.
[13] L. Bertero, C. Barbano, et. al., Unitopatho, 2021. doi:10.21227/9fsv-tm25.
[14] C. A. Barbano, Eidoslab/torchstain: v1.2.0-stable, 2022. URL: https://doi.org/10.5281/zenodo.6979540. doi:10.5281/zenodo.6979540.
[15] Z. Cai, R. C. Poulos, J. Liu, Q. Zhong, Machine learning for multi-omics data integration in cancer, iScience 25 (2022) 103798. URL: https://www.sciencedirect.com/science/article/pii/S2589004222000682. doi:https://doi.org/10.1016/j.isci.2022.103798.
[16] E. Tartaglione, C. A. Barbano, C. Berzovini, M. Calandri, M. Grangetto, Unveiling covid-19 from chest x-ray with deep learning: a hurdles race with small data, International Journal of Environmental Research and Public Health 17 (2020) 6933.
[17] C. A. Barbano, E. Tartaglione, C. Berzovini, M. Calandri, M. Grangetto, A two-step radiologist-like approach for covid-19 computer-aided diagnosis from chest x-ray images, in: ICIAP, 2022.
[18] E. Tartaglione, C. A. Barbano, M. Grangetto, End: Entangling and disentangling deep representations for bias correction, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021, 2021.
[19] R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer statistics, 2021, CA: A Cancer Journal for Clinicians 71 (2021) 7–33. doi:10.3322/caac.21654.
[20] H. A. H. Chaudhry, R. Renzulli, D. Perlo, F. Santinelli, S. Tibaldi, C. Cristiano, M. Grosso, G. Limerutti, A. Fiandrotti, M. Grangetto, et al., Unitochest: A lung image dataset for segmentation of cancerous nodules on ct scans, in: ICIAP, 2022.
[21] H. A. H. Chaudhry, R. Renzulli, D. Perlo, F. Santinelli, S. Tibaldi, C. Cristiano, M. Grosso, A. Fiandrotti, M. Lucenteforte, D. Cavagnino, Lung nodules segmentation with deephealth toolkit, in: ICIAP, 2022.
[22] L. A. anf Alessandro Mazzei, Coronary artery calcium scoring: Its practicality and clinical utility in primary care, in: Progress in Artificial Intelligence - 17th Portuguese Conference on Artificial Intelligence, EPIA 2015, Coimbra, Portugal, September 8-11, 2015. Proceedings, 2015.
[23] H. S. Peter Kamel, Paul Yi, C. T. Lin, Prediction of coronary artery calcium and cardiovascular risk on chest radiographs using deep learning, Radiology: Cardiothoracic Imaging (2021).
[24] H. Peng, et al., Accurate brain age prediction with lightweight deep neural networks, MedIA (2021).
[25] C. Wachinger, et al., Detect and correct bias in multi-site neuroimaging datasets, MedIA (2021).
[26] B. Glocker, et al., Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects, in: MedNeurIPS Workshop, 2019.
[27] B. Dufumier, et al., Openbhb: a large-scale multi-site brain mri data-set for age prediction and debiasing, NeuroImage (2022).
[28] C. A. Barbano, B. Dufumier, E. Duchesnay, M. Grangetto, P. Gori, Contrastive learning for regression in multi-site brain age prediction, 2023.
[29] U. A. Gava, F. D'Agata, E. Tartaglione, R. Renzulli, M. Grangetto, F. Bertolino, A. Santonocito, E. Bennink, G. Vaudano, A. Boghi, M. Bergui, Neural network-derived perfusion maps: a model-free approach to computed tomography perfusion in patients with acute ischemic stroke, Frontiers in Neuroinformatics (2023).
[30] D. Perlo, E. Tartaglione, U. Gava, F. D'Agata, E. Benninck, M. Bergui, Unitobrain dataset: A brain perfusion dataset, in: ICIAP, 2022.
[31] IIGM, Italian Istitute for Genomic Medicine, https://www.iigm.it/site/, 2021.
[32] S. Tarallo, G. Ferrero, G. Gallo, et. al., Altered fecal small rna profiles in colorectal cancer reflect gut microbiome composition in stool samples., mSystems (2019).
[33] A. Thomas, P. Manghi, F. Asnicar, et. al, Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation, Nature Medicine (2019).
[34] S. Pernice, P. Castagno, L. Marcotulli, M. Maule, L. Richiardi, G. Moirano, M. Sereno, F. Cordero, M. Beccuti, Impacts of reopening strategies for covid-19 epidemic: a modeling study in piedmont region, BMC Infectious Diseases (2020).