Artificial Intelligence in Medicine: From Imaging to Omics

Carlo Adornetto[1,*], Pierangela Bruno[1], Francesco Calimeri[1], Edoardo De Rose[1] and Gianluigi Greco[1]

[1]Department of Mathematics and Computer Science, University of Calabria, Via Pietro Bucci, Rende, Italy

Abstract

In the healthcare domain, the analysis of images and omics data by means of methods based on Artificial Intelligence (AI) is increasingly providing meaningful insights into a number of diseases. However, accurate diagnoses and effective treatments require, in general, complex and multimodal analyses of medical data; making fruitful use of such data, due to their complex, sensitive and heterogeneous nature, is not straightforward. In recent years, Machine Learning and Deep Learning techniques emerged as powerful tools to perform specific disease detection and classification. In this work, we describe the ongoing projects of our research group related to this field.

Keywords

Artificial Intelligence, Medical Imaging, Genomics, Deep Learning

1. Introduction

With the advancement of technology and the availability of data, Artificial Intelligence (AI) has become an essential tool for medical professionals to improve patient outcomes, optimize treatment plans, and facilitate the diagnosis of various medical conditions. In particular, AI has been extensively applied in medical imaging and omics analysis, which has led to more accurate diagnoses and personalized treatment plans. Among the different approaches of AI, Deep Learning (DL) has become a popular method for medical image analysis due to its ability to automatically learn relevant features from images and provide accurate results.

Medical imaging is a technique used to create visual representations of the internal anatomy of the human body for clinical analysis and medical intervention. Medical imaging techniques include X-rays, magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), and ultrasound imaging. AI-based techniques have been developed to analyze medical images and provide accurate diagnoses for various medical conditions, such as cancer, cardiovascular diseases, and neurological disorders. Unfortunately, collecting large medical image datasets is typically a non-trivial process. Consequently, it is not always possible to acquire the huge amounts of data needed for effectively tackling more general tasks, such as the diagnosis of several diseases. In this context, novel applications have been proposed to allow models to learn continuously from new streams of data without forgetting the previously learned knowledge. This is the principle of Continual Learning (CL), a novel paradigm by which the trained model is computed via data streams where tasks and data are only available over time. With this approach, the model can incrementally learn and autonomously change its behavior without forgetting the original task [1]. Indeed, in medical contexts, the training from scratch of models that would be able to accomplish and predict a wide number of possible tasks and activities requires a huge (annotated) dataset, containing, for example, all possible disease or image modalities. This makes such approaches non-scalable. In a more realistic scenario, physicians receive a model already trained on typical tasks and, when a (completely) new activity occurs, the model can be trained again to solve new problems [2].

Considering the ability of DL to learn complex patterns, such methods have been particularly successful also in analyzing omics data. Omics analysis involves the study of biological molecules, including DNA, RNA, proteins, and metabolites. They can provide insights into the underlying molecular mechanisms of diseases and help identify biomarkers for diagnosis, prognosis, and personalized treatment. AI-based techniques, particularly DL, have been extensively applied in omics analysis, including genomics, transcriptomics, proteomics, and metabolomics. DL methods have been shown to provide accurate predictions of disease outcomes, discover novel biomarkers, and support clinicians in the proper treatment design. In the field of *functional genomics*, starting from the results of the Human Genome Project, the evolution of *high-throughput* and *next-generation sequencing* techniques provide big volumes of genomic-scale data.

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy

*Corresponding author.

carlo.adornetto@unical.it (C. Adornetto); pierangela.bruno@unical.it (P. Bruno); francesco.calimeri@unical.it (F. Calimeri); edoardo.derose@unical.it (E. D. Rose); gianluigi.greco@unical.it (G. Greco)

0000-0002-9734-1017 (C. Adornetto); 0000-0002-0832-0151 (P. Bruno); 0000-0002-0866-0834 (F. Calimeri); 0000-0002-0032-9434 (E. D. Rose); 0000-0002-5799-6828 (G. Greco)

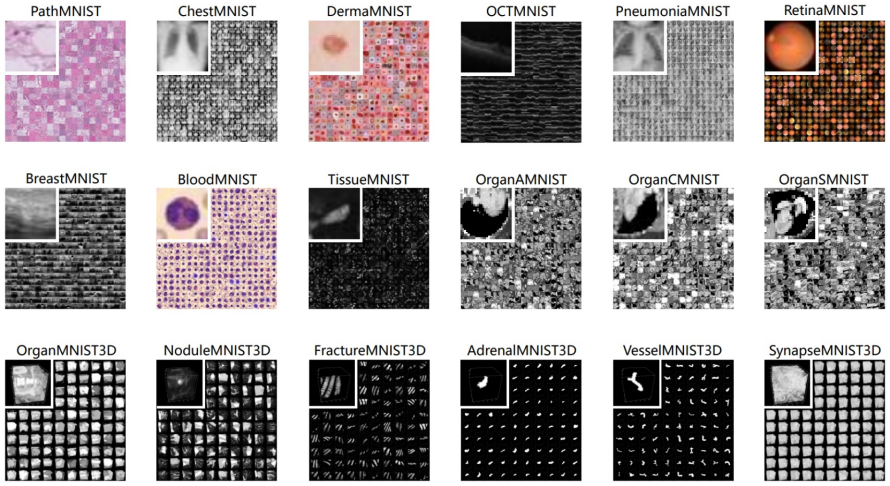


Figure 1: Example of images in MedMNIST dataset. Source: https://medmnist.com/

Indeed, by using such sequencing techniques it is possible to measure the expression of thousands of genes for each patient and hence to collect quantitative Gene Expression Profiles (GEP) to be used for research and clinical purposes. In recent years, DL has been widely adopted in this field, providing breakthrough results and meaningful insights into the relationship between genomics and cancer. A number of recent studies propose and evaluate new approaches for feature selection (FS) on GEP for cancer diagnosis and prognosis[3]. They aim to select the most informative genes able to characterize classes and identify groups of patients.

Although very promising, DL models are in general not immediately interpretable, making it difficult to understand the causal relationship between the inputs and their outcomes. In the bioinformatics domain, this is an even more severe problem. Models interpretability is indeed crucial to understand, for example, in the case of genomics, how the expression of a gene affects the progression of oncological disorders.

In this context, the adoption of Explainable Artificial Intelligence (XAI) methods has started to gain momentum for interpretability purposes as well as to enhance FS [4, 5, 6].

More in general, the black-box nature of Neural Networks (NNs) often limits the interpretability of their results. Advances in XAI provide various methods for interpreting black box models, offering a clearer understanding of their predictions. For example, shapely Additive ex-Planation (SHAP) [7] is a game theory-based approach for interpreting black box models. It assigns an importance value, called a SHAP value, to each feature, based on its contribution to the predictions. The SHAP method provides a way to understand the underlying workings of NNs predictions, leading to improved insights and better decision-making. Another state-of-the-art XAI approach for imaging is Gradient-weighted Class Activation Mapping (Grad-CAM) [8] which is a popular technique for visualizing and understanding decisions made by a CNN model. Specifically, it allows for visualizing the regions of an input image that are most important for predicting a particular class. In particular, Grad-CAM produces a heatmap that highlights the relevant regions in the input image that the model used during the decision-making process.

Despite the increasing number of DL-based applications in medical imaging and omics analysis, there are still challenges that need to be addressed, such as the limited availability of high-quality annotated datasets, the interpretability of DL models, and the generalization of the models to new datasets. Therefore, our research activities aim at facing such challenges by designing innovative solutions based on DL. In the following sections, we describe our research activities applied to medical imaging, in Section 2, and genomics-scale data analysis, in Section 3. Eventually, we discuss the future direction and challenges for AI in the medical field and draw the conclusion in Section 4.

2. AI for Medical Imaging

2.1. Continual Learning for medical image classification

We proposed the use of CL approach to support DL architectures in classifying medical images [2]. In particular, we used a collection of standardized biomedical photos,

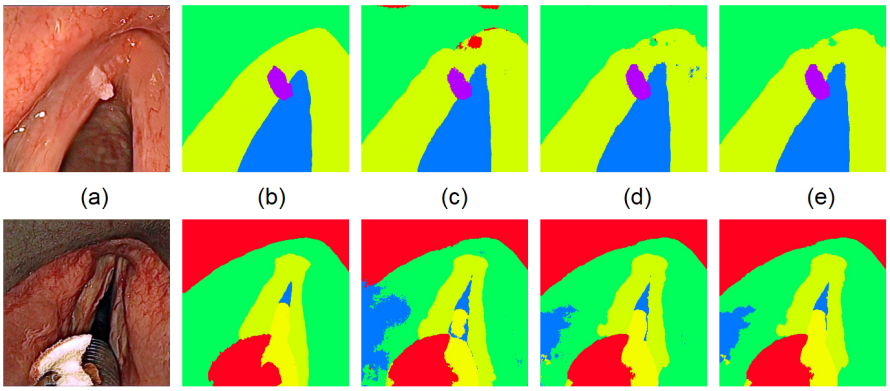


Figure 2: Example results obtained using 2 different patients. From left to right: raw image (a), ground truth segmentation (b), semantic segmentation obtained using the U-net network (c), semantic segmentation obtained includin ASP in the training (d), post-processing applied on the results displayed in **d** (e).

MedMNIST v2 [9] that is composed of 12 datasets for 2D images and 6 datasets for 3D. An example of the dataset is shown in Figure 1.

We focused our experiments on the following data sets: Colon Pathology, Dermatoscope, Retinal OCT, Blood Cell Microscope, and Kidney Cortex Microscope. We used and compared different CL strategies (i.e., Naive, Replay, CWR*, ICaRL, Cumulative) [10].

Our results show that Replay strategies do not reach a good performance; this is probably due to the unbalanced distribution of images across classes and a high number of images.

Instead, we found promising results for both CWR* and ICaRL according to the accuracy and forgetting value.

2.2. XAI in Chest X-ray classification

We investigated the use of CNNs to perform multiple-disease classification from Chest X-ray images [11]. Diseases that are a matter of concern for our experiments are COVID-19 and tuberculosis (TB) Pneumonia. In particular, we achieved promising results in the classification task (i.e., Recall mean value: 0.89).

Furthermore, we analyzed the CNNs-based model to identify the mechanisms and the motivations steering NN decisions in the classification task. Specifically, we used Grad-CAM to identify the image areas used by the model to take a particular decision. The Grad-CAM technique uses the gradient information flowing into the final convolutional layer of a NN to compute the importance of each feature map. It then weights the activation maps of the final convolutional layer based on the importance of each feature map and averages them to produce a heatmap that highlights the important regions in the input image.

Then, the highlighted regions represent the most important features in the classification of a particular disease, as shown in Figure 3.

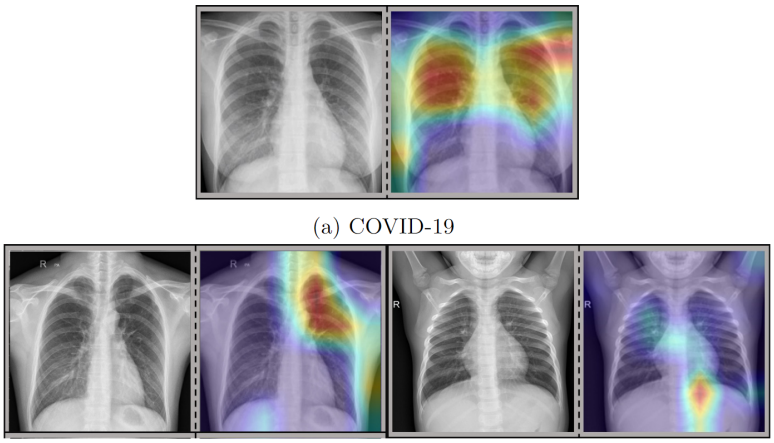
Once these regions are identified, they can be analyzed to discover patterns in Chest X-ray images related to the disease in question.

In addition to identifying patterns, we used these highlighted regions to analyze the correlation between these areas and classification accuracy. This involves examining the performance of the model when these regions are removed and comparing it to the performance when the full image is used. If the removal of these regions results in a significant reduction in classification accuracy, it suggests that these regions are indeed important for accurate disease diagnosis.

In particular, in our experiments, we performed both quantitative and qualitative analyses to confirm that the portions identified by Grad-CAM are actually significant. As for a quantitative assessment, a substantial decrease of Recall after the removal (on average around 5%) is shown. As for a qualitative assessment, we took advantage from the TB dataset, which features labels provided by expert clinicians, representing the approximate location of the abnormality in the lung. In many cases, Grad-CAM highlights the same areas suggested by clinicians.

2.3. Segmentation

Semantic segmentation deals with labeling each pixel on the image with a specific class. In the medical field, semantic segmentation represents a huge breakthrough and it has been widely used for the detection of tumors, identifying different organs, and classifying specific tissue [12].



(a) COVID-19

(b) TB Pneumonia (c) Healthy patients

Figure 3: Visual example of achieved results. For each diagnostic class, we show the raw Chest X-ray image (left) and Grad-CAM result (right). Images on the right sides highlight the most important areas involved in the classification process.

2.3.1. Laryngeal Endoscopic Images

We proposed a DL-based approach to perform semantic segmentation of Laryngeal Endoscopic Images [13, 14]. The dataset consists of 536 manually segmented in vivo color images (512×512 pixels) of the larynx captured during two different resection surgeries. In particular, images are categorized in 7 classes: *void, vocal folds, other tissue, glottal space, pathology, surgical tool and intubation.* In order to improve the quality of our prediction, we took advantage of the potential coming from the declarative nature of rule-based languages such as Answer Set Programming (ASP). In particular, we included ASP in the training to drive NN decisions and penalize misclassification according to prior knowledge. Also, we refined the prediction via rule-based methods, removing noise (i.e., small "islands" of misclassified pixels) and wrong predicted classes (i.e., classes that do not respect medical requirements).

Figures 2 show a visual example of the results achieved by U-Net on two images.

In general, our approach achieved a mean IoU value higher than 0.7, with a relevant improvement after post-processing.

3. AI for Omics Data Analysis

Functional genomics data, and in particular GEP datasets, represent a valuable source of information in medicine: they are indeed used for diagnosis, prevention, and precision medicine. However, their analysis results are challenging for three main reasons. The first one is the *course of dimensionality*: a genomics dataset typically consists of a very large number of genes (features) for a small number of patients (samples); the second problem concerns *imbalanced classes*: in most of the cases, there is a significant difference between the number of instances in each group of interest. Finally, sequencing data are typically collected from multiple sources, different laboratories, and sequencing tools. This results in noisy datasets which are difficult to analyze [15].

We proposed a new algorithm for genomic-scale analysis, based on DL and XAI, whose aim is threefold: first, select the most meaningful genes for a regression/classification problem; second, provide a more accurate prediction model; third, quantify and evaluate the feature's contribution to the predictions through XAI [16]. The proposed algorithm is based on two main ideas: (1) recognize similarly correlated features using clustered correlation matrix and then filter the redundant information for each group by using Autoencoders (AEs). In contrast with previous works, where AEs are used for dimensionality reduction [17], we implemented a mechanism to still work at the level of the original features. We hence provide a more treatable dataset in terms of dimensionality, without affecting interpretability; (2) we train NNs and we iteratively select the most meaningful features using a new ad-hoc defined XAI score. We eventually use the set of selected features (from all the iterations) to train and explain a final model. The proposed algorithm workflow is shown in 4.

We used our algorithm for the GEP analysis of Chronic Lymphocytic Leukemia (CLL) patients. The dataset was composed of the GEP of 97 patients for 19367 genes. For each patient a real-valued number was provided, indicating, as a factor of prognosis, the time interval after which the condition of the patient deteriorates. We distinguished two classes: a first one for patients whose

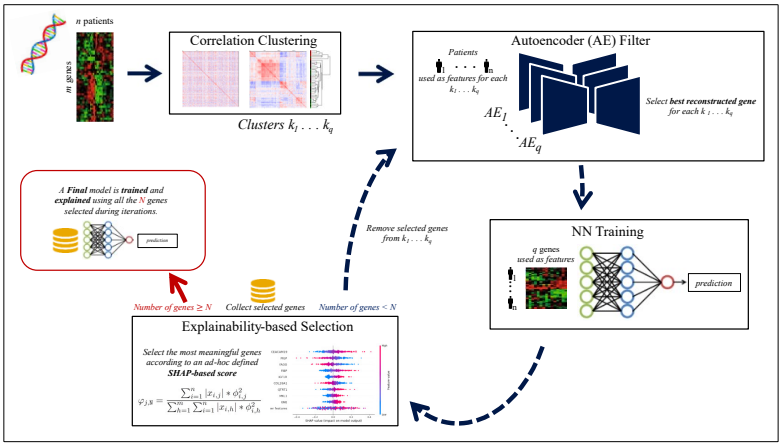
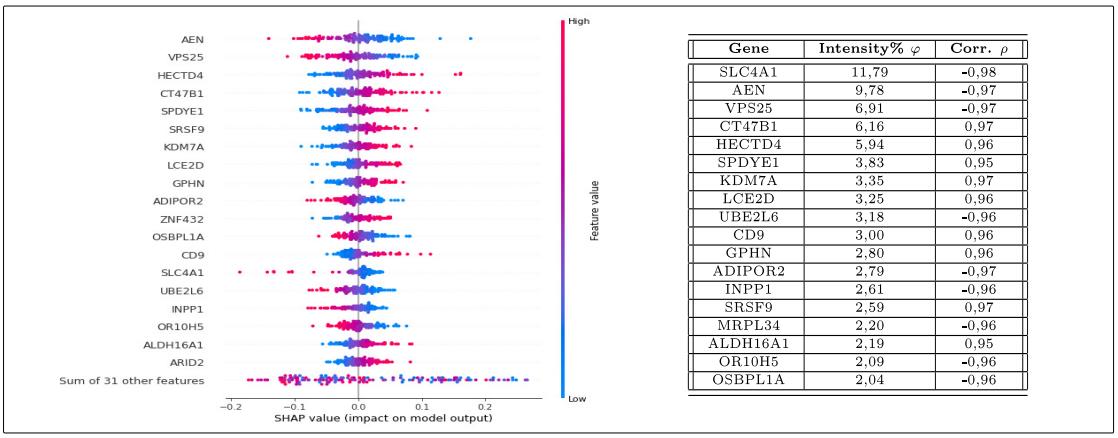


Figure 4: As a preliminary step, the algorithm performs clustering of the correlation matrix to identify q clusters of similarly correlated genes. For each cluster, an AE is trained by using genes as samples and patients as features. The q best-reconstructed genes are then selected as the most representative of each cluster. A NN is trained for prediction using such genes. The ad-hoc defined XAI-score is then used to select the most meaningful genes on the NN predictions. Such genes are removed from the analysis and the algorithm iterate until the number of genes selected by the XAI-score is less than a certain threshold N. A final model is trained and explained by using all the XAI-selected genes over the iterations.



Gene	Intensity% φ	Corr. ρ
SLC4A1	11,79	-0,98
AEN	9,78	-0,97
VPS25	6,91	-0,97
CT47B1	6,16	0,97
HECTD4	5,94	0,96
SPDYE1	3,83	0,95
KDM7A	3,35	0,97
LCE2D	3,25	0,96
UBE2L6	3,18	-0,96
CD9	3,00	0,96
GPHN	2,80	0,96
ADIPOR2	2,79	-0,97
INPP1	2,61	-0,96
SRSF9	2,59	0,97
MRPL34	2,20	-0,96
ALDH16A1	2,19	0,95
OR10H5	2,09	-0,96
OSBPL1A	2,04	-0,96

Figure 5: On the left: beeswarm plot of SHAP values computed on the final model resulting from the algorithm. On the right: the final set of genes selected according to the ad-hoc defined XAI-based score (Intensity) and Correlation

condition deteriorates in a period shorter than 24 months, and a second class for the patients whose condition deteriorates in a period equal to or longer than 24 months. We used the proposed algorithm for training a NN to solve such a classification problem as well as to identify a set of meaningful genes over the whole set of 19367. Such selected genes, together with their ad-hoc defined XAI-score are reported in Figure 5.

In a more general context, the proposed algorithm can be suitably used as a tool in genomics to identify protective (or not) sets of genes for a disease, suggesting potential pathways for further medical investigation. Our preliminary study will evolve into a large-scale testing campaign to assess the algorithm's performance on a wide set of state-of-the-art genomics datasets.

4. Conclusion

AI has been playing an increasingly important role in the medicine and healthcare domain over the past few years, particularly in the areas of medical imaging and omics data analysis. Nevertheless, new techniques and methods are needed for properly facing complex issues, such as scarce or low-quality data, yet improving diagnos-

tic accuracy, disease prediction, and treatment planning, leading to better patient outcomes and improved quality of life. However, the integration of AI into clinical practice must be carefully managed to ensure that these technologies are used ethically, responsibly, and safely. Ongoing research and development in this field will be critical to realizing the full potential of AI in medicine and improving patient care.

Acknowledgements

This work has been partially funded by PON "Ricerca e Innovazione" 2014-2020, CUP: H25F21001230004.

References

[1] C. S. Lee, A. Y. Lee, Clinical applications of continual learning machine learning, The Lancet Digital Health 2 (2020) e279–e281.

[2] A. Quarta, P. Bruno, F. Calimeri, Continual learning for medical image classification, in: F. Calimeri, M. Dragoni, F. Stella (Eds.), 1st AIxIA Workshop on Artificial Intelligence For Healthcare (HC@AIxIA 2022) co-located with the 21st International Conference of the Italian Association for Artificial Intelligence (AIxIA 2022), Udine, Italy, November 30 2022, volume 3307 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 67–76. URL: https://ceur-ws.org/Vol-3307/paper7.pdf.

[3] E. Alhenawi, R. Al-Sayyed, A. Hudaib, S. Mirjalili, Feature selection methods on gene expression microarray data for cancer classification: A systematic review, Computers in Biology and Medicine 140 (2022) 105051.

[4] G. Graham, N. Csicsery, E. Stasiowski, G. Thouvenin, W. H. Mather, M. Ferry, S. Cookson, J. Hasty, Genome-scale transcriptional dynamics and environmental biosensing, Proceedings of the National Academy of Sciences 117 (2020).

[5] J. Meena, Y. Hasija, Application of explainable artificial intelligence in the identification of squamous cell carcinoma biomarkers, Computers in Biology and Medicine 146 (2022) 105505.

[6] M. R. Karim, M. Cochez, O. Beyan, S. Decker, C. Lange, Onconetexplainer: explainable predictions of cancer types based on gene expression data, in: 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), IEEE, 2019, pp. 415–422.

[7] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in neural information processing systems 30 (2017).

[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 618–626.

[9] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, B. Ni, Medmnist v2: A large-scale lightweight benchmark for 2d and 3d biomedical image classification, arXiv preprint arXiv:2110.14795 (2021).

[10] V. Lomonaco, L. Pellegrini, A. Cossu, A. Carta, G. Graffieti, T. L. Hayes, M. De Lange, M. Masana, J. Pomponi, G. M. Van de Ven, et al., Avalanche: an end-to-end library for continual learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3600–3610.

[11] P. Bruno, F. Calimeri, Understanding automatic pneumonia classification using chest x-ray images, in: AIxIA 2020–Advances in Artificial Intelligence: XIXth International Conference of the Italian Association for Artificial Intelligence, Virtual Event, November 25–27, 2020, Revised Selected Papers, Springer, 2021, pp. 37–50.

[12] T. Dhamija, A. Gupta, S. Gupta, R. Katarya, G. Singh, Semantic segmentation in medical images through transfused convolution and transformer networks, Applied Intelligence 53 (2023) 1132–1148.

[13] M.-H. Laves, J. Bicker, L. A. Kahrs, T. Ortmaier, A dataset of laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation, International journal of computer assisted radiology and surgery 14 (2019) 483–492.

[14] P. Bruno, F. Calimeri, C. Marte, M. Manna, Combining deep learning and asp-based models for the semantic segmentation of medical images, in: Rules and Reasoning: 5th International Joint Conference, RuleML+ RR 2021, Leuven, Belgium, September 13–15, 2021, Proceedings 5, Springer, 2021, pp. 95–110.

[15] L. Koumakis, Deep learning models in genomics; are we there yet?, Computational and Structural Biotechnology Journal 18 (2020) 1466–1473.

[16] C. Adornetto, G. Greco, A new deep learning and xai-based algorithm for features selection in genomics, 2023. arXiv:2303.16914.

[17] P. Danaee, R. Ghaeini, D. A. Hendrix, A deep learning approach for cancer detection and relevant gene identification, in: Pacific symposium on biocomputing 2017, World Scientific, 2017, pp. 219–229.