CEUR-WS.org/Vol-4121/Ital-IA_2025_paper_26.pdf

Advances in AI for Health and Medicine @ Computer-Human Interaction Laboratory (CHILab)

Germano Ammirata[1], Salvatore Contino[1,*], Luca Cruciata[1], Irene Siragusa[1] and Roberto Pirrone[1]

[1]*Department of Engineering, University of Palermo, Palermo, 90128, Sicily, Italy*

Abstract

The Computer-Human Machine Interaction Laboratory (CHILab) conducts research on the application of artificial intelligence (AI) and innovative technologies to human health and life sciences. CHILab is focused on multidisciplinary research, with the aim of developing intelligent solutions to emerging challenges in healthcare, biotechnology and pharmaceuticals. Principal research lines spread towards Big Data infrastructures for precision medicine with integration of genomic and chemical data and extension of the DICOM standard for transparent integration of AI models into PACS systems, and distributed models based on Federated Learning. In addition, contributions in Natural Language Processing field include Retrieval Augmented Generation architectures enriched with structured domain knowledge, while explainable AI techniques are investigated and involved for drug discovery and repositioning tasks.

Keywords

Bioinformatics, Medical Imaging, Federated Learning, RAG, Visual Language Models,

1. Introduction

The Computer-Human Interaction Laboratory (CHILab) is a laboratory of the University of Palermo, which carries out research in the application and development of innovative Artificial Intelligence (AI) technologies in the field of human health and life sciences. Up to now, applications of AI in the life sciences are expanding with great consistency and continuity, providing active support for diagnosis. In fact, the laboratory's research activity identifies intelligent processing of medical data as its main research task. This revolution is taking place in the analysis of biomedical data in various fields, from medical imaging to genomics to electronic medical records and drug discovery, thanks to advanced methodologies such as deep learning and natural language processing (NLP) that allow knowledge to be extracted from large masses of complex data (Acosta et al., 2022). In medical image analysis, for instance, deep neural networks (in particular CNNs) can identify lesions or pathologies in X-rays, CT scans or histopathological images with diagnostic accuracy comparable to that of a medical specialist, improving the speed and efficiency of diagnoses [1]. Similarly, in omics data (genomics, proteomics, etc.), machine learning algorithms integrate large-scale genomic datasets to discover genotype-phenotype correlations, identify novel biomarkers and predict the function of genes or variants, contributing to precision medicine [2]. Electronic health records (EHRs) also benefit from AI: NLP techniques applied to unstructured clinical texts enable the automatic extraction of relevant information from medical notes and the prediction of clinical outcomes using neural networks trained on longitudinal patient data (Yang et al., 2022). Finally, in drug discovery, artificial intelligence (e.g. through generative models and deep neural networks) accelerates the discovery of drugs by proposing and evaluating new molecules in silico: deep learning techniques allow virtual screening of vast chemical libraries and predicting the pharmacological and toxicological properties of drug candidates, making the identification of promising molecules more efficient [3]. The research activities of the Human-Machine Interaction laboratory are

---

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy

[*] Corresponding author.

germano.ammirata@unipa.it (G. Ammirata); salvatore.contino01@unipa.it (S. Contino); luca.cruciata@unipa.it (L. Cruciata); irene.siragusa02@unipa.it (I. Siragusa); roberto.pirrone@unipa.it (R. Pirrone)

0000-0002-7476-1545 (S. Contino); 0009-0005-8434-8729 (I. Siragusa); 0000-0001-9453-510X (R. Pirrone)

© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2025-12-07

contributing to the evolution in all these fields, thanks to the joint activities of its various members. Specifically, the next sections will describe the individual activities divided by task, reporting the results already achieved and a description of the activities currently being conducted.

2. DICOM extension for application of AI models @ CHILab

The rigorous regulations and restrictions governing access to medical data, intended to guarantee security and privacy with a view to protecting personal and clinical data, which are considered one of the most sensitive categories, frequently lead healthcare institutions to make such information inaccessible to the external world.

However, the growing attention of the scientific community, driven by the fast advancement of emerging technologies such as artificial intelligence (AI), proposes the development of solutions able to support the diagnostic and prediction phases of the patient's clinical outcome.

Considering the impossibility of transferring clinical data outside their storage locations, it is crucial to find a way to integrate AI models directly within the hospital ecosystem. This insight led to the definition of a new data structure described in the article IODeep [4]. This work introduces an innovative solution that allows the transparent integration of neural networks within the DICOM standard. Through the development of a dedicated DICOM object (IOD), it is possible to store and retrieve both the architecture and the trained weights of a neural network directly within the PACS (Picture Archiving and Communication System). In the manuscript, a real-life use case is illustrated in which the entire diagnostic process performed by a physician using a DICOM Viewer is emulated: starting with the selection of the study up to the analysis of the individual slices of a three-dimensional volume, acquired by imaging techniques such as MRI, CT or PET. The DICOM Viewer is enriched with advanced functionalities that allow the physician to invoke the predictive model capable of automatically identifying and highlighting suspicious areas (ROI, Region of Interest). In order to optimise prediction accuracy, the metadata contained in the DICOM file is extracted and used to select the best model for the image type, anatomical region and resolution. The prediction generated by the model is overlaid upon the original image, although the physician will accept or deny the suggestion by validating or rejecting the proposed ROI. The latter, if validated, will be saved within an RT Structured Set, generated using the information contained in the original DICOM and associated with the model predictions.

3. Federated Learning for distributed AI models @ CHILab

Federated Learning (FL) [5] is a machine learning paradigm that allows several clients to collaborate and train a common model while keeping their data local. Unlike typical centralized techniques, which require aggregating data in a single location, FL allows each device to train on its own data and only communicate model weight updates with a central server. This architecture preserves data privacy by guaranteeing that sensitive information never leaves the client, making it especially useful in contexts such as healthcare, where data confidentiality and regulatory compliance are crucial as observed in the previous section.

FL supports heterogeneity in real-world systems, creating interesting possibilities [6]. However, this diversity presents serious challenges. One of the most critical is data heterogeneity, which occurs when the distribution of data varies significantly between clients. This circumstance is referred to as Non-IID (Non-Independent and Identically Distributed) data, and it may lead to inconsistent model updates, resulting in delayed convergence and reduced model accuracy. To address these limitations, our research focuses on innovative strategies for increasing model resilience and fairness in extremely varied contexts. We are investigating ways that will provide a more balanced and fair representation of the global dataset.

Our Centralized Dataloader technique [7] uses metadata about local datasets to facilitate training data selection across clients. Instead of depending entirely on local sampling, the central server can control which data subsets are used during each training round, thus improving model performance while ensuring privacy and decentralization. By carefully regulating how data effects model updates, we may enhance results for all participating clients, even in circumstances of considerable fluctuation.

While Federated Learning is designed to protect privacy, it may nevertheless pose hazards, such as the leakage of sensitive information through model changes. To address this issue, a variety of privacy-preserving techniques are used to enhance security; for example, some solutions introduce noise into model updates, while others minimize the number of a shared parameters or perform aggregation on encrypted models using methods such as homomorphic encryption.

Looking ahead, our study intends to close the gap between theoretical advances in FL and their application in real-world scenarios. We are working to make Federated Learning a realistic and trustworthy tool for distributed AI in healthcare and beyond by improving client coordination, refining strategies for dealing with data heterogeneity, and strengthening privacy measures.

4. Generative AI with RAG architectures in biomedical domain @ CHILab

In the context of PNRR FAIR Transversal Project 2: “Vision, Language and Multimodal Challenges”[1], for which the focus was on building multimodal data sets and models for domain-specific applications, we proposed MedPix 2.0 [2] [8]. Thanks to the availability of both textual clinical reports and medical scans, such as CT or MR images, the medical domain proved to be a valuable candidate for AI-based multimodal application. MedPix 2.0 data set was built as an appropriate reorganization of MedPix® [3] for its usage in the development of AI models. It consists of a total of 2050 medical images, 1062 CT and 988 MR scans, associated to 671 clinical cases. More specifically, textual information are arranged in JSON documents, thus allowing the access to specific fields, such as the scan typology, the body part shown, or the found disease in the target clinical report.

Using this information, a classification task is designed to determine the typology of the given medical image and the body part shown. Approaches involving Visual Language Models (VLMs) were built, such as DR-Minerva [9], which employs a Retrieval Augmented Generation (RAG) based architecture [10] with decoder-only models such as Minerva Large Langauge Model (LLM) [11] and Flamingo VLM [12, 13]. In doing this, the prompt given in input to DR-Minerva was enriched following an in-context learning strategy: similar clinical cases to the to be classified were provided as a few-shot sample, thus enhancing the classification capabilities of the model.

To further enhance the classification capabilities of DR-Minerva, the proposed architecture was extended with a Knowledge Base (KB), which serves as an external source to suggest medical diagnosis. The KG was automatically built, using the LlamaIndex framework [4] and Llama 3.1 8B Instruct [14] from *ad hoc* derived documents from MedPix 2.0. More specifically, for each clinical report, in addition to the clinical case descriptions, an academic description of the disease found is reported. The complete system combines the output of the first classification (scan typology and body part) with the retrieved context from the KG, to produce a verbose diagnostic suggestion, which is generated from Llama 3.1 8B Instruct.

5. Computer Aided Drug Discovery and Re-purposing via Explainable AI @ CHILab

One of the main research activities is focused in the pharmaceutical research domain and in particular in the Drug Discovery domain. This is the process of identifying and developing new drugs, involving biological targets and optimising compounds as shown in figure ??. Deep learning techniques have

---

[1] https://fondazione-fair.it/en/transversal-projects/tp2-vision-language-and-multimodal-challenges/

[2] https://github.com/CHILab1/MedPix-2.0.git

[3] https://medpix.nlm.nih.gov/home

[4] https://github.com/jerryjliu/llama_index

shown great promise in virtual screening, a computational technique that identifies bioactive lead compounds for specific protein targets. In collaboration with the computational chemistry group of the RiMED Foundation, CHILab has identified protein targets involved with tumour aetiology (e.g. the CDK family) as the main proteins under investigation. In the early years, an intensive study of the most commonly molecular descriptors within the domain was carried out with particular emphasis on Molecular Fingerprints. In the papers [15, 16], the study of molecular descriptors was deepened, testing the various types of Molecular Fingerprints for the Virtual Screening task. Subsequently, in the paper [17] a new 3D descriptor consisting of the 7 best Molecular Fingerprints (EMBER) was presented outperforming the state-of-the-art approaches. In the last years, the activity of the laboratory in the context of the PNRR project Samothrace is working on the development of in silico models for personalised cancer therapy. In this perspective, a Hierarchical Graph Explainer (HGE) is being developed, which aims to improve the performance of screening models by identifying pharmacophoric molecular moieties not directly involved in bioactivity classification.

6. Usage of structured domain knowledge @ CHILab

When dealing with non-standardized data, ambiguity problems arise when the same entity is referred to different terminology. An example can be found in clinical cases in MedPix 2.0, which are freely generated since no external medical taxonomy is used. This degree of freedom results in clinical cases that show differences in some given field while referring to the same entity. In the disease field of some clinical reports, the term `adenocarcinoma of pancreas` is referred to as `pancreatic ductal adenocarcinoma` and `pancreatic adenocarcinoma`, or misspelled `pancreatic adenocarinoma`. A strategy to overcome the observed variability is needed to allow subsequent AI-based classification.

In fact, existing medical taxonomies and ontologies can be used as valuable external sources for developing an automatic annotation pipeline and providing a standardized version of the data set itself. Medical resources such as the International Classification of Diseases 11th Revision [5] (ICD-11), the Human Disease Ontology (DOID) [18] or the LOINC ontology[6], made up of the Logical Observation Identifiers Names and Code (LOINC) and the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT), can be used in this context. We aim to develop an automatic annotation pipeline that can retrieve the most similar standard term given a raw term from the data set, using LLM or similarity metrics such as the BERTScore [19].

In addition, the usage of ontologies can be later involved in developing an Ontology-Agumented Generation (OAG) approach, in which generative capabilities of LLM in providing a diagnostic suggestion, as in DR-Minerva, are directed in a semantic-driven manner, thanks to the ontology, and not with simple RAG.

7. Conclusions and future works

The Computer-Human Interaction laboratory as described in the previous sections is specialised in the creation of innovative technologies in the field of artificial intelligence offering solutions in the biomedical domain. Through the many years, the lab has acquired and developed expertise in the processing of data regardless of the source from which it is obtained. In fact, the skills of the lab members try to cover all aspects of the domain by integrating AI algorithms from molecular data such as molecular representations describing drugs and their interaction with proteins, up to the treatment of complex clinical cases by combining structured and unstructured data through multimodal approaches. Recently, CHILab is actively working on the development of new multimodal analysis paradigms, integrating genomic and pharmaceutical analysis within Big Data infrastructures in the context of precision medicine. Furthermore, it is developing systems that integrate RAGs by integrating medical

---

[5] https://icd.who.int/en/

[6] https://loincsnomed.org/

ontologies in order to make the retrieval of information contextualised. Additionally, in order to make the real-world applications of IODeep possible by expanding its functionality within various hospital facilities, a new Federated Learning approach is being developed that will allow the federated training of models. In this way, it will be possible to improve the performance of local models without moving data from their storage location. Finally, Few-shot segmentation techniques aimed at overcoming the lack of labelled data in the domain of interest are under exploration.

Acknowledgments

This work is supported by the cup project B73C22000810001, project code ECS_00000022 “SAMOTHRACE” (Sicilian MicronanoTech Research And Innovation Center), and the cup project J73C24000070007, "CAESAR" (Cognitive evolution in Ai: Explainable and Self-Aware Robots through multimodal data processing). The works presented were partially developed on the Leonardo supercomputer with the support of CINECA-Italian Super Computing Resource Allocation class C project IscrC_DOC-VLM (HP10C64J82).

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-4o and Quillbot for grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and takes full responsibility for the publication’s content.

References

[1] R. Aggarwal, V. Sounderajah, G. Martin, D. S. W. Ting, A. Karthikesalingam, D. King, H. Ashrafian, A. Darzi, Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis, NPJ Digital Medicine 4 (2021) 65. doi:10.1038/s41746-021-00438-z.

[2] W. S. Alharbi, M. Rashid, A review of deep learning applications in human genomics using next-generation sequencing data, Human Genomics 16 (2022) 26. doi:10.1186/s40246-022-00396-x.

[3] A. V. Sadybekov, V. Katritch, Computational approaches streamlining drug discovery, Nature 616 (2023) 673–685. doi:10.1038/s41586-023-05905-z.

[4] S. Contino, L. Cruciata, O. Gambino, R. Pirrone, Iodeep: An iod for the introduction of deep learning in the dicom standard, Computer Methods and Programs in Biomedicine 248 (2024) 108113. doi:10.1016/j.cmpb.2024.108113.

[5] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y. Arcas, Communication-efficient learning of deep networks from decentralized data, in: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, PMLR, 2017, p. 1273–1282. URL: https://proceedings.mlr.press/v54/mcmahan17a.html.

[6] M. Ye, X. Fang, B. Du, P. Yuen, D. Tao, Heterogeneous federated learning: State-of-the-art and research challenges, ACM Computing Surveys 56 (2023). doi:10.1145/3625558.

[7] G. Ammirata, G. J. Pezzullo, S. Contino, B. D. Martino, R. Pirrone, Federated learning framework for privacy-preserving ai in healthcare, in: L. Barolli (Ed.), Advanced Information Networking and Applications, Springer Nature Switzerland, Cham, 2025, p. 316–325. doi:10.1007/978-3-031-87778-0_31.

[8] I. Siragusa, S. Contino, M. L. Ciura, R. Alicata, R. Pirrone, MedPix 2.0: A Comprehensive Multimodal Biomedical Data set for Advanced AI Applications with Retrieval Augmented Generation and Knowledge Graphs, arXiv preprint arXiv:2407.02994v3 (2025).

[9] I. Siragusa, S. Contino, R. Pirrone, DR-Minerva: A Multimodal Language Model Based on Minerva for Diagnostic Information Retrieval, in: AIxIA 2024 – Advances in Artificial Intelligence, 2025, pp. 288–300.

[10] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al., Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, Advances in Neural Information Processing Systems (2020).

[11] R. Orlando, L. Moroni, P.-L. H. Cabot, E. Barba, S. Conia, S. Orlandini, G. Fiameni, R. Navigli, Minerva LLMs: The First Family of Large Language Models Trained from Scratch on Italian Data, Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024) (2024).

[12] J.-B. Alayrac, J. Donahue, P. Luc, A. Miech, I. Barr, Y. Hasson, K. Lenc, A. Mensch, K. Millican, M. Reynolds, R. Ring, E. Rutherford, S. Cabi, T. Han, Z. Gong, S. Samangooei, M. Monteiro, J. Menick, S. Borgeaud, A. Brock, A. Nematzadeh, S. Sharifzadeh, M. Binkowski, R. Barreira, O. Vinyals, A. Zisserman, K. Simonyan, Flamingo: a Visual Language Model for Few-Shot Learning, arXiv preprint arXiv:2204.14198 (2022).

[13] A. Awadalla, I. Gao, J. Gardner, J. Hessel, Y. Hanafy, W. Zhu, K. Marathe, Y. Bitton, S. Gadre, S. Sagawa, J. Jitsev, S. Kornblith, P. W. Koh, G. Ilharco, M. Wortsman, L. Schmidt, OpenFlamingo: An Open-Source Framework for Training Large Autoregressive Vision-Language Models, arXiv preprint arXiv:2308.01390 (2023).

[14] LlamaTeam, The Llama 3 Herd of Models, arXiv preprint arXiv:2407.21783 (2024).

[15] I. Mendolia, S. Contino, U. Perricone, R. Pirrone, E. Ardizzone, A Convolutional Neural Network for Virtual Screening of Molecular Fingerprints, volume 11751 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2019, p. 399–409. URL: http://link.springer.com/10.1007/978-3-030-30642-7_36. doi:`10.1007/978-3-030-30642-7_36`.

[16] I. Mendolia, S. Contino, U. Perricone, E. Ardizzone, R. Pirrone, Convolutional architectures for virtual screening, BMC Bioinformatics 21 (2020) 310. doi:`10.1186/s12859-020-03645-9`.

[17] Ember—embedding multiple molecular fingerprints for virtual screening 23 (2022) 2156. doi:`10.3390/ijms23042156`.

[18] L. M. Schriml, J. B. Munro, M. Schor, D. Olley, C. McCracken, V. Felix, J. A. Baron, R. Jackson, S. M. Bello, C. Bearer, et al., The human disease ontology 2022 update, Nucleic acids research 50 (2022) D1255–D1261.

[19] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, Y. Artzi, BERTScore: Evaluating Text Generation with BERT, Preprint arXiv:1904.09675, 2020.