

Deep Learning e calcolo ad alte prestazioni per l'elaborazione di immagini biomediche

Marco Aldinucci, Claudio Berzovini, Costantino Grana, Marco Grangetto, Luca Pireddu, Gianluigi Zanetti

Dipartimento di Informatica, Università di Torino

Città della Salute e della Scienza Torino

Dipartimento di Ingegneria "Enzo Ferrari", Università di Modena Reggio Emilia
Data Intensive Computing, CRS4

marco.aldinucci@unito.it, marco.grangetto@unito.it, cberzovini@cittadellasalute.to.it,
costantino.grana@unimore.it, luca.pireddu@crs4.it, gianluigi.zanetti@crs4.it,

Abstract

Il progetto DeepHealth, recentemente finanziato dalla Commissione Europea, ha come obiettivo la realizzazione di un ecosistema europeo costituito da piattaforme di calcolo ad alte prestazioni, librerie software e competenze multi-disciplinari di intelligenza artificiale, calcolo parallelo e scienze mediche per l'elaborazione e la diagnosi basata su immagini. Il contributo presenta sinteticamente le competenze e le infrastrutture nazionali coinvolte nel progetto.

1 Introduzione

Il settore salute rappresenta da sempre una sfida globale di primaria importanza e costituisce un settore chiave da molteplici punti di vista che vanno dagli aspetti puramente medici e scientifici, a quelli tecnologici e dell'innovazione, agli aspetti economici. Questo ha determinato nel corso degli anni imponenti iniziative di supporto della ricerca internazionale e in particolare Europea, volte a risolvere le principali sfide del settore con elevato impatto per il benessere dei cittadini [European Commission, 2012].

Le enormi potenzialità dell'applicazione dei metodi di intelligenza artificiale al settore medico sono evidenti da molti anni [Coiera, 1996], in particolare per quanto riguarda l'utilizzo di sistemi di supporto all'attività clinica e diagnostica basata su sistemi esperti capaci di modellare il ragionamento in presenza di incertezza. Nonostante le grandi aspettative, le maggiori potenzialità dell'intelligenza artificiale in ambito medico sono ancora confinate al mondo della ricerca e non hanno ancora dato il loro potenziale contributo nella pratica clinica a stretto contatto con il paziente.

Il grande successo determinato dall'applicazione delle reti neurali profonde (Deep Learning) per l'apprendimento automatico [LeCun *et al.*, 2015] nei contesti più disparati che vanno dall'elaborazione e comprensione del linguaggio naturale alla visione artificiale stanno generando un impatto, inimmaginabile fino a pochi anni fa, in un numero sterminato di settori dall'elettronica di consumo, all'agricoltura di precisione, alla finanza. In questo contesto sono diventati fattori critici da

un lato la disponibilità di una grande base dati di conoscenza e dall'altro di elevate capacità di calcolo.

Diventano oggi evidenti le enormi opportunità determinate da una maggiore penetrazione delle tecnologie dell'intelligenza artificiale in ambito medico a patto di stimolare progetti interdisciplinari capaci di far incontrare l'innovazione IA da un lato, le competenze mediche dall'altro, supportate da una gestione più consapevole dell'enorme quantità di dati medico diagnostici generati dai sistemi sanitari nazionali. Il progetto *DeepHealth*, "Deep-Learning and HPC to Boost Biomedical Applications for Health", finanziato dalla Commissione Europea nel 2019, costituisce un esempio di questo tipo di approccio in cui il finanziamento pubblico alla ricerca mette intorno allo stesso tavolo esperti di IA, calcolo ad alte prestazioni e personale medico con l'ambizione di innescarne l'imponente potenziale.

Nel presente contributo saranno brevemente presentati il progetto DeepHealth (sezione 2) e le competenze apportate dai partner italiani coinvolti: Università di Torino (sezione 3), Modena Reggio Emilia (sezione 4), Centro di Ricerca Sviluppo e Studi Superiori in Sardegna CRS4 (sezione 5), Città della Salute e della Scienza di Torino (sezione 6).

2 Il progetto DeepHealth

Il progetto DeepHealth è stato finanziato dalla Commissione Europea per un periodo di tre anni a partire dal 2019 attraverso la call H2020 ICT-11-2018-2019- HPC and Big Data enabled Large-scale Test-beds and Applications (Grant Agreement 825111). Il consorzio comprende 21 partners da 9 differenti paesi membri: Spagna, Olanda, Romania, Grecia, Germania, Francia, Italia, Svezia e Svizzera. Il gruppo è fortemente multi-disciplinare, include 9 organizzazioni di ricerca, 4 aziende ospedaliere, 4 grandi industrie e 4 SME che complessivamente coprono le seguenti aree di competenza: Intelligenza Artificiale e Machine Learning, Big-Data, High Performance Computing, sviluppo di applicazioni cliniche e sistemi di supporto alla diagnosi.

In particolare, il progetto DeepHealth è dedicato al settore della diagnosi per immagini. In questo contesto supporta tutti gli aspetti dell'elaborazione di immagini biomediche includendo nel consorzio tutti gli attori a partire dagli utenti finali, agli esperti di elaborazione immagini ai gestori di

centri di super calcolo. DeepHealth permetterà di definire una piattaforma capace di sfruttare in modo trasparente per l'utente sistemi di calcolo ad alte prestazioni di tipo eterogeneo per supportare i modelli di Deep Learning con alti livelli di parallelismo e accelerazione hardware. DeepHealth intende risolvere problemi concreti e supportare con soluzioni innovative i bisogni reali dell'analisi clinica. Per questo motivo il progetto prevede ben 14 casi d'uso di applicazione del Deep Learning all'elaborazione di immagini mediche che verranno sviluppate su 7 differenti piattaforme di calcolo. Gli ambiti clinici coinvolti dal progetto sono molteplici: malattie neurologiche, classificazione e predizione di patologie tumorali, classificazione automatica di immagini istopatologiche. Il progetto promuoverà inoltre la definizione di un approccio standard per la gestione, indicizzazione ed elaborazione delle immagini mediche ai fini dell'apprendimento automatico aprendo la strada a un miglior sfruttamento dell'enorme fonte di conoscenza accumulata in termini di dati clinici ed oggi non ancora opportunamente valorizzata.

3 Università di Torino: HPC4AI e calcolo ad alte prestazioni per l'intelligenza artificiale

L'Università di Torino contribuisce al progetto con competenze multidisciplinari di image processing (<http://di.unito.it/eidoslab>) e calcolo parallelo (<http://alpha.di.unito.it>) del Dipartimento di Informatica e alcuni casi d'uso in collaborazione con i Dipartimenti di Scienze Mediche e Neuroscienze. Il progetto sfrutta la piattaforma di calcolo HPC4AI (<https://hpc4ai.it>), supportata da finanziamenti della Regione Piemonte, sotto la guida del Dipartimento di Informatica dell'Università di Torino. Tale progetto catalizza le competenze presenti nei due principali atenei piemontesi in ambito di Edge/IoT, Machine Learning, HPC, Big Data Analytics in un centro di competenza federato dotato di sistemi di calcolo ad alte prestazioni facilmente fruibili [Aldinucci and others, 2018]. HPC4AI mira a soddisfare la necessità strategica di creare nel contesto regionale e nazionale la capacità di comprendere e anticipare la domanda del mercato nell'ambito soprattutto dell'intelligenza artificiale, creando un ecosistema vitale, duraturo e sostenibile.

4 Università di Modena e Reggio Emilia

L'Università di Modena e Reggio Emilia partecipa al progetto con un gruppo di lavoro composto da ricercatori in ICT del laboratorio AImageLab (<http://aimagelab.unimore.it>) e uno con riconosciute competenze in ambito dermatologico. UNIMORE è responsabile della progettazione e implementazione della European Computer Vision Library (ECVL) che integrerà librerie software già esistenti in un framework comune. L'obiettivo è quello di consentire lo sviluppo con più linguaggi di programmazione (Java, C++, Python) e il supporto di diversi sistemi operativi, tipi di dati e formati di dati per l'imaging scientifico. Nell'ambito del progetto intendiamo ottenere una libreria estremamente compatibile, con la possibilità di riutilizzare tutte le funzionalità delle librerie esistenti e con alcuni dei suoi algoritmi sviluppati per sfruttare le potenzialità di AI distribuito che verranno rese disponibili dal progetto.

5 CRS4

Il CRS4 è un centro multidisciplinare che contribuisce al progetto DeepHealth su più fronti. Dal punto di vista dello sviluppo infrastrutturale, il CRS4 contribuirà a rendere il toolkit DeepHealth facilmente ed efficientemente fruibile su infrastrutture di tipo cloud. Attraverso un approccio di tipo infrastructure-as-code si potranno approvvisionare in maniera automatica ambienti di calcolo distribuito per supportare l'addestramento scalabile di grandi reti neurali utilizzando grosse moli di dati, anche su infrastrutture di calcolo fornite come servizio (i.e., Infrastructure-as-a-Service) [Ruttkies and others, 2018].

Dal punto di vista delle applicazioni del framework obiettivo di DeepHealth, il CRS4 in collaborazione con il Karolinska Institutet svilupperanno una piattaforma per la patologia digitale col supporto per l'annotazione automatica di vetrini. Il CRS4 e il Karolinska Institutet utilizzeranno la piattaforma con un dataset di oltre 13K immagini di vetrini di biopsie di tumore alla prostata per creare un'applicazione per assistere i patologi nelle attività di diagnosi.

6 Città della Salute

La Città della Salute e della Scienza di Torino è la maggiore azienda Ospedaliera del Piemonte e una delle più grandi d'Italia; presso la S.C. Radiologia 2, dotata di 4 apparecchiature TC (Tomografia Computerizzata) vengono eseguiti annualmente circa 42000 esami, il 50% circa per patologie oncologiche. Nell'ambito del progetto la Radiologia 2 si occupa di classificazione e predizione delle malattie oncologiche per l'analisi dei noduli polmonari. L'obiettivo, nell'ambito del progetto DeepHealth, è di formare un team multidisciplinare in grado di creare una raccolta di esami TC contenenti immagini di noduli polmonari, opportunamente archiviati e anonimizzati, su cui eseguire un lavoro di segmentazione e annotazione del nodulo. Il dataset così creato sarà utilizzato come caso d'uso per il toolkit DeepHealth in modo da valutare il potenziale beneficio in termini di miglioramento della performance diagnostica.

Riferimenti bibliografici

- [Aldinucci and others, 2018] Marco Aldinucci et al. HPC4AI, an AI-on-demand federated platform endeavour. In *ACM Computing Frontiers*, Ischia, Italy, May 2018.
- [Coiera, 1996] Enrico W Coiera. Artificial intelligence in medicine: the challenges ahead. *Journal of the American Medical Informatics Association*, 3(6):363–366, 1996.
- [European Commission, 2012] European Commission. Redesigning health in europe for 2020. http://www.e-health-com.eu/fileadmin/user_upload/dateien/Downloads/redesigning_health-eu-for2020-ehf-report2012_01.pdf, 2012.
- [LeCun et al., 2015] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- [Ruttkies and others, 2018] Christoph Ruttkies et al. PhenomeNal: Processing and analysis of Metabolomics data in the Cloud. 12 2018.