

**UNIVERSITÀ DEGLI STUDI DI MILANO**

selezione pubblica per n.1 posto/i di Ricercatore a tempo determinato ai sensi dell'art.24, comma 3, lettera a) della Legge 240/2010 per il settore concorsuale 01/B1 - Informatica, settore scientifico-disciplinare INF/01 - Informatica presso il Dipartimento di Informatica Giovanni Degli Antoni, (avviso bando pubblicato sulla G.U. n. 46 del 11/06/2021) Codice concorso 4759

## Alessandro Petrini

### CURRICULUM VITAE

**INFORMAZIONI PERSONALI**

COGNOME	PETRINI
NOME	ALESSANDRO
DATA DI NASCITA	25/07/1981

**TITOLI****Titoli di studio**

- Dottorato di ricerca in Informatica, conseguito il 22/03/2021 presso il dipartimento di Informatica Giovanni Degli Antoni, Università degli Studi di Milano. Titolo della tesi: "High Performance Computing Machine Learning Methods for Precision Medicine". Relatori: prof. Giorgio Valentini e prof. Giuliano Grossi
- Laurea magistrale di secondo livello in Informatica, conseguita il 19/10/2017 presso il dipartimento di Informatica Giovanni Degli Antoni, Università degli Studi di Milano. Titolo della tesi: "Un algoritmo parallelo su GPU per la stima del moto inter-frame nel codec WebM-VP8". Voto: 110/110 summa cum laudae. Relatori: prof. Giuliano Grossi e prof. Federico Pedersini
- Laurea triennale di primo livello in Matematica per le Applicazioni conseguita il 22/10/2014 presso il dipartimento di Matematica Federico Enriques, Università degli Studi di Milano. Titolo della tesi: "Algoritmi paralleli per il calcolo di PageRank". Voto: 100/110. Relatori: prof. Luca Pavarino e prof. Paolo Boldi
- Maturità scientifica conseguita presso L.S.S. Leonardo da Vinci nell'anno 2000

**POSIZIONI**

01/12/2020 - in corso	2021 Assegnista (tipo B) presso il dipartimento di Informatica dell'Università degli Studi di Milano
06-23/05/18 01-15/12/19 15-18/02/20	Visiting researcher presso Kircher Lab, Berlin Institute of Health (von Humboldt University)
01/10/2017 - 22/03/2021	Dottorando presso Università degli Studi di Milano - dipartimento di Informatica
01/11/2018 - 31/07/2020	Contratto di Collaborazione Continuativa presso Università Vita-Salute San Raffaele
01/09/2017 - 31/08/2018	Contratto di Collaborazione Continuativa presso Ospedale San Raffaele
19/02/2015 - 31/12/2016	Collaboratore di ricerca presso Università degli Studi di Milano - dipartimento di Informatica (incarico di prestazione d'opera di natura intellettuale)

## ATTIVITA' DI RICERCA

La sua principale attività di ricerca si colloca nel campo del calcolo parallelo e accelerato, con applicazioni in svariati ambiti scientifici. Ha sviluppato tecniche di calcolo accelerato e parallelo per risolvere problemi nei campi della bioinformatica, apprendimento automatico, elaborazione delle immagini, compressione di segnali video e algoritmica su grafi.

Una schematizzazione delle principali linee di ricerca su cui il candidato ha concentrato i suoi sforzi è riportata in seguito, seguita dalla descrizione dettagliata di ogni linea.

- 1) CALCOLO PARALLELO E SVILUPPO DI METODI DI MACHINE LEARNING PER HIGH PERFORMANCE COMPUTING (HPC)
  - a) Studio di metodi per il calcolo parallelo / accelerato / efficiente e tecnologie associate
  - b) Progettazione e sviluppo per piattaforme HPC di metodi di machine learning per l'analisi di big data
- 2) BIOINFORMATICA
  - a) Sviluppo di metodi di apprendimento massivamente paralleli per l'individuazione di varianti patogeniche rare nelle aree non codificanti del DNA
  - b) Sviluppo di metodi di apprendimento automatico accelerati su GPU per la caratterizzazione dell'attività delle regioni regolatorie nel genoma umano
  - c) Sviluppo di metodi accelerati su GPU basati su reti di Hopfield per la risoluzione del problema della predizione della funzione proteica
- 3) ALGORITMI SU GRAFI
  - a) Sviluppo di metodi innovativi per la risoluzione del problema della colorazione di grafi e loro implementazione parallela per GPU
  - b) Sviluppo di metodi accelerati su GPU per il calcolo di PageRank
- 4) ELABORAZIONE DI SEGNALI VIDEO
  - a) Sviluppo di un innovativo sistema di accelerazione per la codifica di segnali video nello standard VP8
- 5) ELABORAZIONE DELLE IMMAGINI
  - a) sviluppo di metodi accelerati su GPU per il bilanciamento automatico dei colori nelle immagini, con applicazioni al bioimaging
- 6) ALGORITMI DI COMPRESSIONE
  - a) Sviluppo di metodi di compressione per reti neurali

## CALCOLO PARALLELO E SVILUPPO DI METODI DI MACHINE LEARNING PER HIGH PERFORMANCE COMPUTING (HPC)

A partire dal corso di laurea triennale il candidato ha dedicato la sua attenzione allo studio delle tecniche e tecnologie legate al calcolo parallelo ed accelerato, trasformando un interesse in una passione e professione. Le attività di specializzazione e di ricerca in questo ambito sono continuate per tutta la durata della laurea magistrale e il dottorato di ricerca, come mostrato dalle due tesi di laurea e dalla tesi di dottorato. Durante la sua formazione e la successiva attività di ricerca il candidato ha infatti acquisito esperienza su tecniche di parallelizzazione e accelerazione su diversi livelli, a partire da quelle che riguardano l'ottimizzazione sul singolo processore (tecniche SIMD e istruzioni vettoriali), multi-processo e multi-processore, fino ad arrivare alla programmazione di periferiche di accelerazione (GPU) e programmazione di super-computer (massively large high performance computing).

Il candidato ha utilizzato le tecniche acquisite per contribuire a vari settori della ricerca scientifica, in principio aumentando l'efficienza di algoritmi e strategie esistenti, e poi sviluppando approcci scientifici innovativi. In particolare, negli ultimi anni il candidato si è concentrato nello studio e sviluppo di tecniche di apprendimento automatico "HPC-aware", ovvero disegnate fin dal principio per sfruttare al meglio le attuali tecnologie di calcolo parallelo e

accelerato.

## BIOINFORMATICA

Nel corso del suo dottorato di ricerca ha affrontato problemi fondamentali e tuttora aperti nel campo della genomica e medicina di precisione, proponendo soluzioni allo stato dell'arte.

In particolare, l'identificazione di varianti genetiche rare patogeniche o deleterie nelle aree non codificanti del DNA è uno dei problemi fondamentali nella medicina di precisione, reso particolarmente difficile per la notevole dimensione dei dati e l'elevato sbilanciamento tra il numero di campioni che rappresentano varianti patogeniche e il numero di campioni di varianti neutre - circa 1:35000. La soluzione proposta e sviluppata dal candidato è un "ensemble di ensemble" altamente parallelo - *parSMURF* [11] - che migliora l'attuale stato dell'arte attraverso il tuning automatico dei parametri di learning, reso possibile da una efficiente implementazione che sfrutta le caratteristiche degli attuali supercomputer per distribuire la computazione su un numero elevato di unità di elaborazione.

I risultati promettenti ottenuti da *parSMURF* hanno suggerito la sua estensione come oggetto di un progetto europeo, di cui il candidato è principale proponente e P.I., premiato dal consorzio PRACE (Partnership for Advanced Computing in Europe) nella 21ma call for projects. Il Progetto ha ottenuto 50 milioni di ore di calcolo sul supercalcolatore SuperMUC-NG presso il Leibniz-Rechenzentrum (Garching, Germania) e ha titolo: "ParBigMen: ParSMURF application to Big genomic and epigenomic data for the detection of pathogenic variants in Mendelian diseases". In dettaglio, ParBigMen ha come obiettivo lo sviluppo di tecniche efficienti per la selezione di variabili epigenomiche, da usare per aumentare la informatività della rappresentazione delle varianti attualmente disponibili, e lo sviluppo di *ParSMURF-NG*, un'implementazione scalabile e con maggiori performance predittive rispetto a *ParSMURF*. Il fine ultimo è quello di usare *ParSMURF-NG* per computare nuovi score predittivi di patogenicità Mendeliana da inserire in *Genomiser* (<https://hpo.jax.org/app/tools/genomiser>), un framework di analisi che effettua il ranking delle varianti regolatorie in base alla loro patogenicità rispetto a specifiche malattie mendeliane.

Un altro problema aperto nell'ambito della medicina di precisione è l'identificazione della attività delle regioni regolatorie nel DNA. Nel contesto del progetto MIUR-DAAD dal titolo: "Developing machine learning methods for the prioritization of regulatory variants in human disease" in collaborazione con il Berlin Institute of Health (von Humboldt University), sono stati proposti in [4,18] dei metodi unimodali basati su Deep Neural Network in grado di competere con l'attuale stato dell'arte. L'obiettivo a lungo termine che il candidato sta al momento affrontando è lo sviluppo di tecniche multi-task e multi-modali per la creazione di un vero e proprio "atlante" delle regioni regolatorie, contestualmente al loro stato di attivazione in diverse linee cellulari e tessuti.

Il problema della predizione della funzione proteica e genica nella Gene Ontology (GO) può essere modellizzato come un problema semi-supervisionato di etichettatura di grafi parzialmente etichettati. Anche in questo problema, lo sbilanciamento tra geni annotati e non-annotati rende il problema difficilmente affrontabile da tecniche di apprendimento automatico tradizionali. Questo problema è stato efficacemente risolto da *COS-Net*, un metodo allo stato dell'arte che sfrutta metodi "cost sensitive" basati su reti di Hopfield. Dal momento che il costo computazionale del metodo lo rende inapplicabile su recenti basi-di-dati con elevata cardinalità, il candidato ha sviluppato e proposto in [16] *Par-COSNet*, una versione accelerata su GPU del pacchetto *COS-Net*, in grado di apportare sostanziali migliorie ai tempi di calcolo rispetto alla versione originale (fino a 350x), senza intaccare la capacità predittiva del metodo. Il pacchetto *Par-COSNet* è stato rilasciato liberamente come software open source ed è tuttora a disposizione della comunità scientifica.

Nell'ambito delle attività del laboratorio di biologia computazionale e bioinformatica AnacletoLab, il candidato ha contribuito anche ad altri progetti rilevanti, come lo sviluppo di metodi gerarchici per la predizione della funzione proteica nella Gene Ontology [5], lo sviluppo di metodi semi-supervisionati per la predizione del fenotipo e prognosi di pazienti [8,29] e la

creazione di tool visuali per l'integrazione di reti biomolecolari sbilanciate.

A partire da Febbraio 2020 è stato inoltre invitato a fare parte dei consorzi internazionali N3C (<https://covid.cd2h.org>) e COVIRT ([www.cov-irt.org](http://www.cov-irt.org)), al fine di contribuire con il disegno e sviluppo di modelli efficienti e innovativi per la diagnosi e prognosi di pazienti COVID-19. A tal fine, sta al momento collaborando nello sviluppo di efficienti metodi per l'imputazione dei dati mancanti, la selezione di variabili e lo sviluppo di metodi trasduttivi basati su grafi [8].

#### ALGORITMI SU GRAFI

Durante il dottorato, il candidato ha collaborato con diversi membri del laboratorio PhuseLab del dipartimento di Informatica dell'Università degli Studi di Milano nella ricerca e sviluppo di un innovativo metodo accelerato su GPU per risolvere il problema della colorazione di grafi. In questo problema, dato un grafo di input, si vuole trovare una suddivisione dei nodi i cui gruppi siano massimamente indipendenti. Altrimenti detto, lo scopo è l'identificazione di gruppi massimali di nodi che non hanno connessioni dirette in comune. In [6,19] il candidato ha proposto un metodo Monte Carlo basato su Catene di Markov, la cui forza risiede nel determinare una soluzione in cui le classi di nodi siano (in probabilità) bilanciate, ovvero formate dallo stesso numero di nodi. L'implementazione su GPU proposta si dimostra essere efficace sia per tempi di calcolo, sia per qualità di soluzione, misurata come scostamento dal valore atteso del numero di nodi per ciascuna classe. L'implementazione proposta è stata rilasciata liberamente come software open source ed è tuttora a disposizione della comunità scientifica.

Molti problemi in svariati campi della ricerca scientifica possono essere modellati come un problema di ranking (ordinamento per importanza) dei nodi di un grafo, e l'individuazione delle pagine internet più importanti - centrali - nel grafo del web è un classico esempio. Nella sua tesi di laurea triennale il candidato ha proposto una metodologia alternativa per il calcolo dell'indice PageRank, l'algoritmo usato originariamente da Google per determinare il ranking dei nodi del grafo del World Wide Web e ordinare i risultati di una ricerca. Il metodo è basato sulla trasformazione del problema originale in un sistema lineare e risolverlo con metodi iterativi. Il candidato ha proposto una implementazione accelerata su GPU che, anche grazie a metodi di rappresentazione compressa dei dati, è risultata competitiva per qualità di soluzione, ma allo stato dell'arte per tempi di calcolo.

#### ELABORAZIONE DEI SEGNALI VIDEO

Nell'ambito del progetto europeo "T-NOVA: Network Functions As-a-Service Over Virtualised Infrastructures", in qualità di collaboratore alla unità Unimi - dip. di Informatica (resp. prof. Giuliano Grossi) e Italtel, il candidato ha realizzato, e presentato in [15, 23, 24], una funzione di rete virtuale per la transcodifica del segnale video. Tale funzione prevede la conversione in tempo reale di un flusso di segnale video tra standard di codifica diversi. Il contributo innovativo di tale funzione riguarda l'utilizzo di GPU per accelerare la codifica nello standard VP8 di Google, e tale contributo è stato integrato nel codificatore software di riferimento, ottenendo una accelerazione del processo di codifica fino a 4 volte.

#### ELABORAZIONE DELLE IMMAGINI

Il candidato sta attualmente collaborando con il laboratorio MIPS del dipartimento di Informatica dell'Università degli Studi di Milano nella realizzazione di algoritmi accelerati su GPU per il bilanciamento automatico del colore e del contrasto di immagini. Tali algoritmi fanno parte della famiglia "Retinex" e l'ulteriore contributo scientifico del candidato riguarda la creazione di algoritmi facenti parti della stessa famiglia, ma atti ad operare su immagini di natura biomedica, come radiografie, e volumi 3D e 4D ottenute attraverso la tecnica diagnostica della risonanza magnetica.

Nell'ambito di una collaborazione con l'ospedale Vita-Salute San Raffaele di Milano, ha realizzato un sistema basato su reti neurali per lo scalpig automatico - identificazione automatica dell'area

cerebrale - di volumi 3D e 4D contenenti risonanze magnetiche di feti in utero.

#### ALGORITMI DI COMPRESSIONE

Nell'ambito del progetto "Multicriteria Data Structures and Algorithms: from compressed to learned indexes, and beyond" finanziato dal PRIN, sta realizzando delle tecniche di compressione di reti neurali atte a ridurre la rappresentazione in memoria di un modello allenato, senza sacrificarne le capacità predittive. Le quattro tecniche sviluppate sono state testate su reti usate per risolvere problemi di classificazione e regressione, mostrando risultati allo stato dell'arte per livelli di compressione raggiunta; è stato inoltre mostrato che le performance predittive della rete non sono inferiori alla controparte non compressa. Infine, grazie a tecniche di ottimizzazione software, è stato mostrato come i tempi di calcolo usati per allenare il modello compresso siano comparabili alla controparte non compressa.

#### GRUPPI E PROGETTI DI RICERCA

**Partecipazione alle attività di un gruppo di ricerca caratterizzato da collaborazioni a livello nazionale o internazionale**

- a) È da più di cinque anni un membro attivo del laboratorio di bioinformatica e biologia computazionale AnacletoLab, Dipartimento di Informatica, Università degli Studi di Milano, presso cui ha preso il dottorato di ricerca in Informatica, sotto la guida del responsabile scientifico Prof. Giorgio Valentini. Le principali attività di ricerca all'interno del laboratorio riguardano lo studio e sviluppo di metodi di apprendimento automatico fortemente paralleli per risolvere problemi di genomica e bioinformatica. In particolare, ha sviluppato metodi di machine learning per la predizione di varianti genomiche rare patogeniche e deleterie in aree non codificanti del DNA, metodi di Deep Learning per la caratterizzazione delle regioni regolatorie e metodi paralleli per la predizione della funzione proteica in reti complesse.
- b) Al presente collabora con gruppi del dipartimento di informatica al fine di portare a termine le seguenti ricerche:
  - a. Studio e sviluppo di nuovi metodi per la rappresentazione efficiente e compressa di dati, nel contesto del progetto finanziato dal PRIN "Multicriteria Data Structures and Algorithms: from compressed to learned indexes, and beyond" (resp. unità Unimi: prof. Marco Frasca);
  - b. Studio e sviluppo di metodi paralleli per il processing di segnali video in collaborazione con il laboratorio di elaborazione delle immagini MIPS Lab.
- c) Collabora con il Jackson Laboratory for Genomic Medicine, CT, USA, sotto la supervisione del prof. Peter Robinson nello sviluppo di metodi accelerati per l'identificazione di varianti patogeniche e deleterie
- d) Ha collaborato con il Kircher Lab presso il Berlin Institute of Health (von Humboldt University) per lo sviluppo di metodi per la caratterizzazione dell'attività delle regioni regolatorie, nell'ambito del progetto finanziato dal MIUR "Developing Machine Learning Methods for the Prioritization of Regulatory Variants in Human Disease".
- e) Ha collaborato con il reparto di neuroradiologia dell'ospedale Vita-Salute San Raffaele (resp. scientifico prof. Andrea Falini) per lo sviluppo di metodi di gestione e archiviazione di segnali RM; in collaborazione con tale gruppo di ricerca ha inoltre sviluppato metodi di automazione per il pre-processamento e analisi di segnali RM.
- f) Ha partecipato al progetto europeo "T-NOVA: Network Functions As-a-Service Over Virtualised Infrastructures" in qualità di collaboratore dell'unità del Dipartimento di Informatica dell'Università degli Studi di Milano (resp. Giuliano Grossi). In tale ambito ha sviluppato un metodo parallelo per la compressione di segnali video e un metodo di

mapping per i servizi di rete virtuale su una infrastruttura di rete.

**Responsabilità scientifica (Principal Investigator – PI) di progetti di ricerca internazionali e nazionali, ammessi al finanziamento sulla base di bandi competitivi che prevedano la revisione tra pari**

- 01 ottobre 2020 - 31 ottobre 2021. Co-Responsabile del progetto Prace 21st call for projects: Proposal 2020225452 - “ParBigMen: ParSMURF application to Big genomic and epigenomic data for the detection of pathogenic variants in Mendelian diseases”. Questo Progetto è stato premiato da “Partnership for Advanced Computing in Europe” con 50 milioni di ore di calcolo sul supercalcolatore SuperMUC-NG presso il Leibniz-Rechenzentrum (Garching, Germania). Questa straordinaria risorsa di calcolo è utilizzata per identificare la patogenicità di varianti genomiche rare in aree non codificanti del DNA attraverso l’applicazione di tecniche di apprendimento automatico fortemente parallele che ha sviluppato nel corso del suo dottorato.
- Giugno 2020 - Agosto 2020: Responsabile del progetto Prace preparatory call type A: Proposal 2010PA5332 - “ParStoBig2: ParSMURF Scaling to Big data”. Partnership for Advanced Computing in Europe. In questo progetto ha verificato le prestazioni del software sviluppato per il progetto “ParBigMen” in termini di scalabilità ed efficienza parallela.

**Partecipazione a progetti di ricerca internazionali e nazionali, ammessi al finanziamento sulla base di bandi competitivi che prevedano la revisione tra pari**

- Novembre 2020 - in corso: Collaboratore al progetto "Multicriteria Data Structures and Algorithms: from compressed to learned indexes, and beyond", PI: Ferragina Paolo, PRIN no. 2017WR7SHH.
- Ottobre 2019: Collaboratore al progetto Prace preparatory call type A: Proposal 2010PA5046 - “ParStoBig: ParSMURF Scaling to Big data”. Partnership for Advanced Computing in Europe
- 1 Gennaio 2018 - 31 Dicembre 2019: Collaboratore al progetto MIUR-DAAD Joint Mobility Program, "Developing Machine Learning Methods for the Prioritization of Regulatory Variants in Human Disease", resp. Prof. Giorgio Valentini e Prof. Martin Kircher, Ministero dell’Istruzione, dell’Università e Ricerca - Deutscher Akademischer Austauschdienst.
- 1 Novembre 2018 - 30 Giugno 2020: Collaboratore presso Università Vita-Salute San Raffaele (Milano) per il progetto “Ottimizzazione reti sistema informatico e archiviazione segnali RM”, resp. Dott. Andrea Falini.
- 1 Gennaio 2018 - 31 Dicembre 2018: Collaboratore al progetto "HPC-SoMuC: Development of Innovative HPC Methods for the Detection of Somatic Mutations in Cancer", finanziato da Cineca e Regione Lombardia tramite LISA 2017-2018 Grant Programs.
- 1 Settembre 2017 - 30 Settembre 2018: Collaboratore presso Ospedale Vita-Salute San Raffaele (Milano) per il progetto “Ottimizzazione reti sistema informatico e archiviazione segnali RM”, resp. Dott. Andrea Falini.
- 1 Gennaio 2017 - 31 Dicembre 2017: Collaboratore al progetto "HyperGeV: Detection of Deleterious Genetic Variations through Hyperensemble Methods.", finanziato da Cineca e Regione Lombardia tramite LISA 2016-2017 Grant Programs.
- 1 Giugno 2015 - 31 Dicembre 2016: Collaboratore al progetto Europeo "T-NOVA: Network Functions As-a-Service Over Virtualized Infrastructures" finanziato dalla comunità europea tramite 7th Framework Programme, Grant Agreement no.

619520. Responsabile unità Università degli Studi di Milano: prof. Giuliano Grossi.

### ATTIVITA' DI SERVIZIO PER LA COMUNITA' SCIENTIFICA

Attività di revisione per riviste internazionali e Partecipazione al comitato scientifico di programma di conferenze internazionali.

È revisore di articoli scientifici per la rivista peer-reviewed Scientific Report. È inoltre nel Program Committee delle conferenze internazionali:

- The 35th ACM Symposium on Applied Computing (SAC 2020)  
Video Processing for Human Behavioral Analysis Track

Progettazione e sviluppo di librerie software rese disponibili alla comunità scientifica in repository pubbliche.

- *parSMURF-NG*  
A parallel and highly scalable machine learning tool for the identification of pathogenic and deleterious SNV - NEW version  
<https://github.com/Topopiccione/parSMURF-NG>
- *MCMC colorer*  
A novel GPU accelerated MCMC algorithm for graph coloring  
[https://github.com/Topopiccione/MCMC\\_Colorer](https://github.com/Topopiccione/MCMC_Colorer)
- *parSMURF*  
A parallel and highly scalable machine learning tool for the identification of pathogenic and deleterious SNV  
<https://github.com/AnacletoLAB/parSMURF>
- *CosNET- GPU (parCOSNet)*  
<https://github.com/AnacletoLAB/ParCOSNet>
- *GPU-accelerated VP8 encoder - for T-NOVA*  
GPU acceleration integrated into the official Google libvpx library  
<https://github.com/Topopiccione/libvpx>
- *Italtel vTU (Virtual transcoding unit) - for T-NOVA*  
<https://github.com/Topopiccione/vTU>
- *TeNOR - Service Mapper module - for T-NOVA*  
<https://github.com/T-NOVA/TeNOR>

### ATTIVITA' COME RELATORE E VISITE DI RICERCA

Formale attribuzione di incarichi di insegnamento o di ricerca (fellowship) presso qualificati atenei e istituti di ricerca esteri o sovranazionali

- Visiting researcher al Berlin Institute of Health, von Humboldt University, per attività di ricerca comuni nell'ambito della Medicina Genomica (Maggio 2018 / Dicembre 2019 / Febbraio 2020)

### DOCUMENTATA ATTIVITA' DI FORMAZIONE PRESSO QUALIFICATI ISTITUTI ITALIANI E STRANIERI

- Partecipazione a "SGP Graduate School on Geometry Processing", 6-7 luglio 2019, Università degli Studi di Milano - Milano
- Partecipazione corso "Debugging and Optimization of Scientific Applications", 19-21 novembre 2018, Cineca - Casalecchio del Reno (BO), Italia
- Partecipazione "2nd International Summer School on Deep Learning", 23-27 luglio 2018, Genova, Italia
- Partecipazione al corso "Deep Dive: parallel programming and performance optimization for KNL architecture", corso on-line con attestato, Colfax Research,

marzo 2017

### ATTIVITA' DIDATTICA

#### Anno Accademico 2020/2021:

- Università degli studi di Milano - Dipartimento di Informatica  
Docente corso extracurricolare "Introduzione a C++" nell'ambito del corso di "Online Game Design" (Resp. Davide Gadia). Corso di 20 Ore (luglio 2021)
- Università degli studi di Milano - Dipartimento di Scienze e Politiche Ambientali  
Tutoraggio e assistenza agli esami per "Statistica e informatica - Modulo Informatica" (Resp. Beatrice Palano)  
Secondo semestre. Durata incarico: 30 ore
- Università degli studi di Milano - Dipartimento di Fisica  
Tutoraggio e assistenza agli esami per "Informatica" (Resp. Carlo Mereghetti)  
Primo semestre. Durata incarico: 45 ore

#### Anno Accademico 2019/2020

- Università degli studi di Milano - Dipartimento di Scienze e Politiche Ambientali  
Tutoraggio e assistenza agli esami per "Statistica e informatica - Modulo Informatica" (Resp. Beatrice Palano)  
Secondo semestre. Durata incarico: 30 ore
- Università degli studi di Milano - Dipartimento di Fisica  
Tutoraggio e assistenza agli esami per "Informatica" (Resp. Carlo Mereghetti)  
Primo semestre. Durata incarico: 45 ore

#### Anno Accademico 2018/2019:

- Università degli studi di Milano - Dipartimento di Informatica  
Docente corso extracurricolare "Introduzione a C++" nell'ambito del corso di "Online Game Design" (Resp. Davide Gadia). Corso di 20 Ore (settembre 2019)
- Università degli studi di Milano - Dipartimento di Scienze e Politiche Ambientali  
Tutoraggio e assistenza agli esami per "Statistica e informatica - Modulo Informatica" (Resp. Beatrice Palano)  
Secondo semestre. Durata incarico: 30 ore
- Università degli studi di Milano - Dipartimento di Fisica  
Tutoraggio e assistenza agli esami per "Informatica" (Resp. Carlo Mereghetti)  
Primo semestre. Durata incarico: 45 ore

#### Anno Accademico 2018/2019:

- Università degli studi di Milano - Dipartimento di Informatica  
Assistenza agli esami per "Linguaggi formali e automi" (Resp. Beatrice Palano)  
Secondo semestre. Durata incarico: 16 ore
- Università degli studi di Milano - Dipartimento di Fisica  
Tutoraggio e assistenza agli esami per "Informatica" (Resp. Carlo Mereghetti)  
Primo semestre. Durata incarico: 45 ore

#### Anno Accademico 2017/2018

- Università degli studi di Milano - Dipartimento di Informatica  
Docente corso extracurricolare "Introduzione a C++" nell'ambito del corso di "Online Game Design" (Resp. Dario Maggiorini). Corso di 20 Ore (settembre 2017)

#### Anno Accademico 2016/2017

- Università degli studi di Milano - Dipartimento di Informatica  
Docente corso extracurricolare "Introduzione a C++" nell'ambito del corso di



“Online Game Design” (Resp. Dario Maggiorini). Corso di 20 Ore (settembre 2016)

Relatore o co-relatore tesi di laurea triennale o magistrale

- Vincenzo Conversano, "Sviluppo di un sistema multimodale per l'analisi automatica di tac polmonari, dati clinici e dati di laboratorio", Comunicazione Digitale BSc, Università degli Studi di Milano, 2021,
- Marco Cavallari, "Modelli di previsione del rischio di mortalità per pazienti COVID", BSc Informatica, Università degli Studi di Milano, 2021,
- Grimi Riccardo, "Multimodal Deep Neural Networks for the prediction of regulatory regions in human genome", Computer Science MS, Università degli Studi di Milano, 2021
- Luca Cappelletti, "Deep Learning methods for the genome-wide tissue specific prediction of active regulatory regions", Computer Science MS, Università degli Studi di Milano, 2019
- Aspes Nicholas, "Parallelization of MCMC methods for graph coloring", Computer Science MS, Università degli Studi di Milano, 2019
- Ghidotti Davide, Computer Science MS, Università degli Studi di Milano, 2019

## **PRODUZIONE SCIENTIFICA**

Pubblicazione e Indici di produttività scientifica

Di seguito vengono riportati gli indici di produttività estratti in data 12/7/2021 dalla banca dati Scopus e da Google Scholar:

Indici	Scopus	Google Scholar
Numero di citazioni	147	261
Indice H	5	5
Indice i10		5

Di seguito sono riportate le riviste scientifiche indicizzate in Scopus e Web of Science in cui ha pubblicato almeno un articolo, con i relativi indici CiteScore (Scopus) e Impact Factor (Web of Science):

Journal name	CiteScore (Scopus)	IF (Web of Science)
Genome Biology	16.5	13.583
Bioinformatics	9.9	6.937
Gigascience	9.1	6.524
Human Brain Mapping	8.4	5.038
IEEE Transactions on Network and Service Management	7.9	4.195
Scientific Report	7.1	4.379
Pattern Recognition Letters	6.7	3.756
BMC Bioinformatics	5.7	3.169
IEEE Access	4.8	3.367
Journal of Real-Time Image Processing	3.5	2.358

A seguito viene riportata la lista completa degli articoli in fase di revisione, gli articoli su riviste

internazionali, i lavori a conferenze internazionali e quelli a conferenze nazionali

### Articoli in fase di revisione in riviste internazionali

- [1] A. Scarabelli, M. Zilocchi, E. Casiraghi, P. Fasani, G. G. Plensich, A. A. Esposito, E. Stellato, **A. Petrini**, G. Valentini, and G. Carrafiello, “Abdominal computed tomography imaging findings in hospitalized covid-19 patients: a year-long experience and associations revealed by explainable artificial intelligence”, *Journal of Computer Assisted Tomography*, Jun 2021. Submitted for review.
- [2] G. G. Plensich, M. Zilocchi, P. Fasani, A. Scarabelli, E. Casiraghi, A. A. Esposito, E. Stellato, G. Valentini, **A. Petrini**, and G. Carrafiello, “Abdominal imaging findings in hospitalized covid-19 patients: a year-long experience,” *Japanese Journal of Radiology*, Jun 2021. Submitted for review.
- [3] S. F. Cappa, F. Conca, E. Catricalà, M. Canini, **A. Petrini**, G. Vigliocco, and P. Della Rosa, “In search of different categories of abstract concepts: a fmri adaptation study”, *Scientific Report*, Jun 2021. Submitted for review
- [4] L. Cappelletti, **A. Petrini**, J. Gliozzo, E. Casiraghi, M. Schubach, M. Kircher, and G. Valentini, “Boosting tissue-specific predictions of active cis-regulatory regions through deep learning and Bayesian optimization techniques”, *BMC Bioinformatics Supplements*, Feb 2021. . Submitted for review.

### Articoli in riviste internazionali

- [5] M. Notaro, M. Frasca, **A. Petrini**, J. Gliozzo, P. N. Robinson, and G. Valentini, “Hemdag: a family of modular and scalable hierarchical ensemble methods to improve gene ontology term prediction,” *Bioinformatics*, vol. 1, Mar 2021. Accepted for publication.
- [6] D. Conte, G. Grossi, R. Lanzarotti, J. Lin, and **A. Petrini**, “Analysis of a parallel MCMC algorithm for graph coloring with nearly uniform balancing”, *Pattern Recognition Letter*, vol. 149, pp. 30–36, 2020
- [7] V. Pieri, F. Sanvito, M. Riva, **A. Petrini**, P. M. V. Rancoita, S. Cirillo, A. Iadanza, L. Bello, A. Castellano, and A. Falini, “Along-tract statistics of neurite orientation dispersion and density imaging diffusion metrics to enhance mr tractography quantitative analysis in healthy controls and in patients with brain tumors”, *Human Brain Mapping*, vol. 42, no. 5, pp. 1268-1286, 2021
- [8] J. Gliozzo, P. Perlasca, M. Mesiti, E. Casiraghi, V. Vallacchi, E. Vergani, M. Frasca, G. Grossi, **A. Petrini**, M. Re, A. Paccanaro, and G. Valentini, “Network modeling of patients’ biomolecular profiles for clinical phenotype / outcome prediction”, *Scientific Report*, vol. 10, no. 1, p. 3612, 2020
- [9] B. R. Barricelli, E. Casiraghi, J. Gliozzo, **A. Petrini**, and S. Valtolina, “Human digital twin for fitness management”, *IEEE Access*, vol. 8, no. 1, pp. 26637-26664, Jan 2020
- [10] M. Canini, P. Cavoretto, P. Scifo, M. Pozzoni, **A. Petrini**, A. Iadanza, S. Pontesilli, R. Scotti, M. Candiani, A. Falini, C. Baldoli, and P. A. Della Rosa, “Subcortico-cortical functional connectivity in the fetal brain: A cognitive development blueprint”, *Cerebral Cortex Communications*, vol. 1, 04 2020
- [11] **A. Petrini**, M. Mesiti, M. Schubach, M. Frasca, D. Danis, M. Re, G. Grossi, L. Cappelletti, T. Castrignanò, P. N. Robinson, and G. Valentini, “parSMURF, a high-performance computing tool for the genome-wide detection of pathogenic variants”, *GigaScience*, vol. 9, 05 2020. giaa052
- [12] N. Zhou, Y. Jiang, ... **A. Petrini**, ..., and I. Friedberg, “The cafa challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens”, *Genome Biology*, vol. 20, no. 1, p. 244, 2019

- [13] L. Cappelletti, J. Gliozzo, **A. Petrini**, and G. Valentini, "Training neural networks with balanced mini-batch to improve the prediction of pathogenic genomic variants in mendelian diseases," *Sensors and Transducers*, vol. 234, no. 6, pp. 16-21, 2019
- [14] P. Perlasca, M. Frasca, C. T. Ba, M. Notaro, **A. Petrini**, E. Casiraghi, G. Grossi, J. Gliozzo, G. Valentini, and M. Mesiti, "Unipred-web: a web tool for the integration and visualization of biomolecular networks for protein function prediction", *BMC Bioinformatics*, vol. 20, no. 1, p. 422, 2019
- [15] G. Grossi, P. Paglierani, F. Pedersini, and **A. Petrini**, "Enhanced multicore-manycore interaction in high-performance video encoding", *Journal of Real-Time Image Processing*, Nov 2018
- [16] M. Frasca, G. Grossi, J. Gliozzo, M. Mesiti, M. Notaro, P. Perlasca, **A. Petrini**, and G. Valentini, "A gpu-based algorithm for fast node label learning in large and unbalanced biomolecular networks", *BMC Bioinformatics*, vol. 19, p. 353, Oct 2018
- [17] M. Kourtis, M. Mcgrath, G. Gardikis, G. Xilouris, V. Riccobene, P. Papadimitriou, E. Trouva, F. Liberati, M. Trubian, J. Batalle, H. Koumaras, D. Dietrich, A. Ramos, J. Riera, J. Bonnet, A. Pietrabissa, A. Ceselli, and **A. Petrini**, "T-nova: An open-source mano stack for nfv infrastructures", in *IEEE Transactions on Network and Service Management*, vol. PP, pp. 1-1, 07 2017

## Proceedings di conferenze internazionali

- [18] L. Cappelletti, **A. Petrini**, J. Gliozzo, E. Casiraghi, M. Schubach, M. Kircher, and G. Valentini, "Bayesian optimization improves tissue specific prediction of active regulatory regions with deep neural networks", in *Bioinformatics and Biomedical Engineering* (I. Rojas, O. Valenzuela, F. Rojas, L. J. Herrera, and F. Ortuno, eds.), (Cham), pp. 600-612, Springer International Publishing, Apr 2020. IWBBIO 2020 International Work-Conference on Bioinformatics and Biomedical Engineering, Granada, Spain
- [19] D. Conte, G. Grossi, R. Lanzarotti, J. Lin, and **A. Petrini**, "A parallel mcmc algorithm for the balanced graph coloring problem", in *IAPR International workshop on Graph-Based Representation in Pattern Recognition*, Tours, France, Jul 2019.
- [20] M. Frasca, M. Sepehri, **A. Petrini**, G. Grossi, and G. Valentini, "Committee-based active learning to select negative examples for predicting protein functions", in *Computational Intelligence Methods for Bioinformatics and Biostatistics* (M. Raposo, P. Ribeiro, S. Serio, A. Staiano, and A. Ciaramella, eds.), (Cham), pp. 80-87, Springer International Publishing, 2020
- [21] C. T. Ba, E. Casiraghi, M. Frasca, J. Gliozzo, G. Grossi, M. Mesiti, M. Notaro, P. Perlasca, **A. Petrini**, M. Re, and G. Valentini, "A graphical tool for the exploration and visual analysis of biomolecular networks," in *CIBB 2018, Computational Intelligence methods for Bioinformatics and Biostatistics*, Caparica, Portugal, Sep 2018
- [22] M. Frasca, M. Sepehri, **A. Petrini**, G. Grossi, and G. Valentini, "Committee-based active learning to select negative examples for predicting protein functions", in *CIBB 2018, Computational Intelligence methods for Bioinformatics and Biostatistics*, Caparica, Portugal, Sep 2018
- [23] P. Comi, P. S. Crosta, M. Beccari, P. Paglierani, G. Grossi, F. Pedersini, and **A. Petrini**, "Hardware-accelerated high-resolution video coding in virtual network functions," in *2016 European Conference on Networks and Communications (EuCNC)*, pp. 32-36, June 2016
- [24] P. Paglierani, G. Grossi, F. Pedersini, and **A. Petrini**, "Gpu-based vp8 encoding: Performance in native and virtualized environments", in *2016 International Conference on Telecommunications and Multimedia (TEMU)*, pp. 1-5, July 2016
- [25] J. F. Riera, J. Batallé, J. Bonnet, M. Dias, M. McGrath, G. Petralia, F. Liberati, A. Giuseppe, A.

Pietrabissa, A. Ceselli, **A. Petrini**, M. Trubian, P. Papadimitrou, D. Dietrich, A. Ramos, J. Melià, G. Xilouris, A. Kourtis, T. Kourtis, and E. K. Markakis, "Tenor: Steps towards an orchestration platform for multi-pop nfv deployment", in *2016 IEEE NetSoft Conference and Workshops (NetSoft)*, pp. 243-250, June 2016.

### Proceedings di conferenze nazionali

[26] **A. Petrini**, M. Schubach, M. Re, M. Frasca, M. Mesiti, G. Grossi, T. Castrignanò, P. Robinson, and G. Valentini, "Parameters tuning boosts hypersmurf predictions of rare deleterious non-coding genetic variants", in *NETTAB 2017, Methods, tools and platforms for Personalized Medicine in the Big Data Era*, Palermo, Italy, Oct 2017

[27] **A. Petrini**, M. Notaro, J. Gliozzo, G. Valentini, G. Grossi, and M. Frasca, "Speeding up node label learning in unbalanced biomolecular networks through a parallel and sparse gpu-based hopfield model," in *BITS 2017, Bioinformatics Italian Society Meeting*, Cagliari, Italy, 2017

[28] P. Perlasca, M. Mesiti, M. Notaro, **A. Petrini**, J. Gliozzo, G. Valentini, and M. Frasca, "A web graphical tool for the integration of unbalanced biomolecular network", in *BITS 2017, Bioinformatics Italian Society Meeting*, Cagliari, Italy, 2017

[29] J. Gliozzo, M. Notaro, **A. Petrini**, P. Perlasca, M. Mesiti, E. Casiraghi, M. Frasca, G. Grossi, M. Re, A. Paccanaro, and G. Valentini, "Modeling biomolecular profiles in a graph-structured sample space for clinical outcome prediction with melanoma and ovarian cancer patients," in *BITS 2017, Bioinformatics Italian Society Meeting*, Cagliari, Italy, 2017

Data

12/07/2021

Luogo

Milano (MI)