



**MISSION
INNOVATION**

accelerating the clean energy revolution

POA MATERIALI AVANZATI PER L'ENERGIA

**PROGETTO IEMAP - Piattaforma Italiana Accelerata per i Materiali per
l'Energia**

D1.24 - Materiali ad alte prestazioni per catodi di accumulatori basati su ossidi metallici

Francesco Buonocore, Marco Catillo, Serena D'Onofrio, Sergio Ferlito,
Simone Giusepponi, Massimo Celino



AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE,
L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

D1.24 - MATERIALI AD ALTE PRESTAZIONI PER CATODI DI ACCUMULATORI BASATI SU OSSIDI METALLICI

Francesco Buonocore (ENEA), Marco Catillo (ENEA), Sergio Ferlito (ENEA), Simone Giusepponi (ENEA),
Massimo Celino (ENEA)

Dicembre 2024

Report MISSION INNOVATION

Ministero dell'Ambiente e della Sicurezza Energetica - ENEA
Mission Innovation 2021-2024 - III annualità

Progetto: Piattaforma accelerata per i Materiali per l'Energia
Work package: IEMAP: Italian Energy Materials Acceleration Platform
Linea di attività LA 1.21: Utilizzo della piattaforma per accumulatori
Responsabile del Progetto: Massimo Celino (ENEA)
Responsabile della LA: Francesco Buonocore (ENEA)

Indice

1	Risultati attesi	5
2	Risultati ottenuti	5
2.1	Avanzamento della ricerca rispetto allo stato dell'arte internazionale con riferimento ai risultati ottenuti.....	5
3	Prodotti attesi.....	6
4	Prodotti ottenuti.....	6
5	Analisi degli scostamenti su attività e risultati	6
6	Sintesi delle attività svolte.....	6
7	Dettaglio delle attività svolte	7
7.1	Introduzione	7
7.2	Architetture informatiche	7
7.3	Metodo computazionale	9
7.4	Dataset.....	10
8	Contributo delle eventuali consulenze alle attività sopra descritte.....	12
9	Pubblicazioni scientifiche	12
10	Eventi di disseminazione	12
11	Descrizione dei risultati ottenuti	13
11.1	Introduzione	13
11.2	Descrizione dei metodi usati	14
11.2.1	<i>Architetture informatiche</i>	14
11.2.2	<i>Metodo computazionale</i>	16
11.2.3	<i>Dataset</i>	17
11.3	Risultati numerici addestramento	17
11.4	Risultati numerici dell'inferenza.....	18
11.5	Confronto con calcoli ab initio.....	21
11.6	Algoritmi predittivi per misure sperimentali.....	24
11.7	Conclusioni	27
11.8	Riferimenti bibliografici	28
11.9	Abbreviazioni ed acronimi.....	31

Indice delle figure

Figura 1. Statistiche dal set di dati Materials Project (MP). (a) Conteggio degli elementi unici racchiusi nelle strutture su base logaritmica (asse y). (b) Frequenza degli elementi nel dataset su base logaritmica. (c) Energia media di formazione per ciascun elemento. Gli elementi di colore grigio non sono presenti nel set di dati MP.	11
Figura 2. Grafici ternari delle proprietà dei materiali: (a) energia di formazione e (b) potenziale redox, previsto dal modello NN; c) soluzione multiobiettivo. Sugli spigoli è riportato il numero di atomi. Scala cromatica: in (a) la concentrazione ottimale è l'area blu, viceversa in (b) e (c) è l'area gialla.....	21
Figura 3. Confronto dell'energia di formazione prevista (cerchi) e calcolata (croci) per sistemi non rilassati. Per i sistemi $\text{Na}_{12}\text{Mn}_3\text{Ni}_6\text{Ti}_3\text{O}_{24}$, $\text{Na}_{12}\text{Mn}_6\text{Ni}_3\text{Ti}_3\text{O}_{24}$, $\text{Na}_{12}\text{Mn}_4\text{Ni}_4\text{Ti}_4\text{O}_{24}$ e $\text{Na}_{12}\text{Mn}_3\text{Ni}_3\text{Ti}_6\text{O}_{24}$, l'incertezza statistica delle energie di formazione previste è indicata da linee continue nere centrate nei valori medi e limitate dalle deviazioni standard.	23
Figura 4. Grafici a dispersione delle energie di formazione previste rispetto a quelle calcolate E_f utilizzando sistemi non rilassati (pannello (a) cerchi rossi) e rilassati (pannello (b) cerchi blu). La linea tratteggiata nera rappresenta la relazione ideale $y = x$	23

Indice delle tabelle

Tabella 1. MAE delle proprietà energetiche di formazione.	18
Tabella 2. L'errore assoluto medio (MAE), il coefficiente di determinazione (R^2) e il coefficiente di correlazione di Pearson (r) sono stati calcolati confrontando le energie di formazione previste e calcolate per i sistemi non rilassati e rilassati.	24

1 Risultati attesi

In questa linea di attività si è previsto di raggiungere i seguenti risultati:

- Applicazione del workflow di automatizzazione dei calcoli ab initio ai database dei materiali esistenti (NOMAD, Materials Project, ...).
- Database ENEA di dati ab initio di materiali catodici basati su ossidi metallici lamellari puri e drogati integrata per accumulatori.
- Analisi machine learning dei materiali cristallini per catodi di batterie e identificazione di potenziali candidati per la sintesi dei materiali.

2 Risultati ottenuti

2.1 Avanzamento della ricerca rispetto allo stato dell'arte internazionale con riferimento ai risultati ottenuti

Nella presente attività sono stati raggiunti i seguenti risultati:

- Applicazione del workflow di automatizzazione dei calcoli ab initio ai database dei materiali esistenti.

I calcoli basati sulla teoria del funzionale densità (DFT) sono stati automatizzati e integrati in una catena di flusso di lavoro computazionale sviluppata all'interno del framework AiiDA (D1.23) applicato a strutture cristalline estratte dal database Materials Project. L'energia totale dei materiali è stata utilizzata per valutare l'energia di formazione e il potenziale di intercalazione. La piattaforma implementata nell'approccio flessibile, scalabile e altamente accurato nella gestione dei workflow e delle simulazioni basato su utilizzo del framework AiiDA, integrato con il cluster CRESCO di ENEA e il codice Quantum Espresso per calcoli DFT, è una tecnologia in linea con lo stato dell'arte (si veda ad esempio T. Lombardo et al., Chem. Rev. 2022, 122, 10899–10969).

- Database ENEA di dati ab initio di materiali catodici basati su ossidi metallici lamellari puri e drogati integrata per accumulatori.

Il set di dati generato in questo studio è disponibile all'indirizzo <https://iemap.enea.it/>. Il workflow AiiDA ha generato fino ad oggi un totale di oltre 700 elaborazioni completate, che sono state memorizzate nel database. Questi dati rappresentano una risorsa fondamentale per l'analisi e l'interpretazione dei risultati. Il database specifico per ossidi metallici lamellari (NaTMO₂) drogati è un avanzamento rispetto allo stato dell'arte (si veda altri database come Materials Project, NOMAD e Materials Cloud).

- Analisi machine learning (ML) dei materiali cristallini per catodi di batterie e identificazione di potenziali candidati per la sintesi dei materiali.

Abbiamo adottato un approccio data-driven in cui la rete neurale a grafo cristallino geometrico (GeoCGNN) è stata utilizzata per prevedere l'energia di formazione come proprietà target dei cristalli generati dopo l'addestramento su un set di dati estratto dal database Materials Project. Abbiamo convalidato la robustezza e la trasferibilità del modello pre-addestrato, offrendo approfondimenti sulla sua più ampia applicabilità e sui suoi limiti quando applicato a diversi set di dati. Abbiamo così dimostrato che la rete GeoCGNN è in grado di prevedere risultati coerenti con i calcoli DFT. Infine, abbiamo reso disponibile un codice che esegua l'analisi con algoritmi ML di un dataset che esplori le variazioni chimico-strutturali e di prestazioni delle batterie in funzione di quantità relative di metalli di transizione negli ossidi misti a struttura laminare, una novità rispetto allo stato dell'arte.

I risultati della presente attività portano vantaggi significativi in termini di:

- sostenibilità, in quanto la rete neurale adottata è in grado di indicare quali materiali sostenibili scegliere per la sintesi dei materiali catodici contribuendo così a ridurre l'impatto ambientale delle batterie elettriche;
- sicurezza, in quanto le batterie agli ioni sodio offrono una maggiore sicurezza rispetto alle batterie al litio, ad esempio, riducendo il rischio di incendi o esplosioni;
- efficienza energetica, in quanto il modello prevede quali siano le composizioni dei materiali catodici con elevati potenziali di cella, il che si traduce in una maggiore efficienza energetica delle batterie agli ioni di sodio.

Questi progressi apporteranno benefici sia al sistema elettrico nazionale sia agli utenti finali.

3 Prodotti attesi

Database ENEA di dati ab initio di materiali catodici basati su ossidi metallici lamellari puri e drogati integrata per accumulatori. Si è previsto il Rapporto tecnico D1.24: "Materiali ad alte prestazioni per catodi di accumulatori basati su ossidi metallici"

4 Prodotti ottenuti

Il database ENEA di dati ab initio di materiali catodici basati su ossidi metallici lamellari puri e drogati integrata per accumulatori è disponibile all'indirizzo <https://iemap.enea.it/>.

Il codice utilizzato per questa attività è disponibile all'indirizzo

<https://gitlab.brindisi.enea.it/claudio.ronchetti/ai4mat>.

Si è emesso il presente Rapporto tecnico D1.24: "Materiali ad alte prestazioni per catodi di accumulatori basati su ossidi metallici"

5 Analisi degli scostamenti su attività e risultati

L'attività si è svolta così come prevista. Non sono state rilevate criticità.

6 Sintesi delle attività svolte

Nella presente attività LA1.21 si è predisposto il workflow messo a punto nella LA1.20 per esplorare un'ampia gamma di composizioni di materiali catodici per batterie agli ioni di Na mediante calcoli di principi primi utilizzando catene di flusso di lavoro. Le strutture cristalline sono state caricate dal database dei materiali Materials Project e quindi eseguendo tramite il workflow automatizzato dei calcoli ab initio e applicando due reti neurali a grafo (GNN). Entrambe le reti sono addestrate su un set di dati open source calcolati con metodi ab initio per prevedere l'energia di formazione che permette di valutare la stabilità dei materiali e il potenziale redox. Questo approccio consente di rilevare l'optimum più velocemente rispetto agli approcci computazionali basati sulla fisica. I dati prodotti sono stati registrati sul database creato e gestito da ENEA. Il codice sviluppato è stato reso disponibile pubblicamente.

7 Dettaglio delle attività svolte

7.1 Introduzione

Nei prossimi paragrafi 7.2, 7.3 e 7.4 saranno illustrate le metodologie adottate, e quindi saranno delineate le architetture delle reti neurali basate sulla teoria dei grafi, l'approccio computazionale ab initio e il dataset da cui si è partiti per addestrare le reti. Nel paragrafo 11 (Descrizione dei risultati ottenuti) i seguenti risultati saranno riportati e discussi dettagliatamente: risultati numerici dell'addestramento; risultati numerici dell'inferenza; confronto tra le predizioni e i calcoli ab initio; diverse nature degli algoritmi predittivi addestrabili partendo da un dataset di misure sperimentali. Inoltre, saranno ripetuti i paragrafi 7.2, 7.3 e 7.4, ma questa volta con i riferimenti di letteratura che non è stato possibile inserire qui per i limiti al numero dei caratteri.

7.2 Architetture informatiche

Nella presente attività si è lavorato su due architetture di rete neurale, CGCNN e GeoCGNN. Queste reti sono state scelte perché, in primo luogo, la CGCNN è un punto di riferimento per la scienza dei materiali e si è consolidata negli ultimi anni, e in secondo luogo era necessaria una versione potenziata che non richiedesse tempi di gestione aggiuntivi (GeoCGNN). Entrambi sono basati su architetture GNN. Le reti neurali e le differenze tra loro sono correlate alla rappresentazione del grafo cristallino e agli strati (layers) della rete neurale.

Rappresentazione del grafo cristallino. Molte ricerche di chimica informatica hanno dimostrato l'importanza critica dei descrittori molecolari, che sono noti per influenzare l'accuratezza del modello più della scelta dell'algoritmo di apprendimento automatico. Per CGCNN, ogni nodo i del grafo cristallino G è caratterizzato da nove vettori di caratteristiche v_i , ciascuno per una proprietà atomica in cui le proprietà continue sono suddivise in dieci categorie, mentre le proprietà discrete possono ovviamente essere codificate automaticamente. Allo stesso modo, ogni spigolo $(i, j)_k$ è rappresentato da un vettore di caratteristiche $u_{(i,j)_k}$, corrispondente all'interazione del legame k -esimo tra i nodi i e j , ed è costruito in modo analogo al descrittore Property-Labeled Material Fragments (PLMF). I PLMF si differenziano per le proprietà di riferimento locali, che includono proprietà generali, misurate e derivate. In primo luogo, la struttura cristallina viene analizzata alla ricerca di vicini atomici tramite tassellatura di Voronoi e quindi il grafo periodico risultante viene scomposto in semplici sottografi. Ci sono diversi sottografi da considerare in generale, ma sono stati limitati a un massimo di tre archi consecutivi e non ripetitivi nel sottografo al fine di frenare la complessità del vettore descrittore finale. Analogamente a CGCNN, GeoCGNN definisce i vicini di ogni atomo all'interno di un raggio specifico risultando in un grafo periodico e definisce le rappresentazioni dei nodi e degli spigoli sul grafo. I nodi sono rappresentati come vettore di caratteristiche one-hot che dipende dal numero atomico, mentre gli archi sono rappresentati come vettore di distanza tra gli atomi (grafo diretto). Inoltre, GeoCGNN utilizza un descrittore che registra il vettore reticolare e il volume della cella per mantenere le informazioni globali.

Rete neurale. CGCNN e GeoCGNN immettono direttamente la struttura atomica e la convertono in un grafo, dove gli atomi sono interpretati come nodi e i legami interatomici sono interpretati come spigoli. Principalmente, entrambe le reti si basano su due insiemi di strati (layers): il primo insieme di strati apprende le caratteristiche comuni a tutte le proprietà del materiale e l'ultimo insieme di strati è separato in più capi per apprendere le caratteristiche specifiche di ciascuna proprietà del materiale. Gli strati convoluzionali del grafo vengono utilizzati per estrarre caratteristiche comuni rilevanti da interazioni di vicini a coppie e, attraverso

più strati, rappresentano anche interazioni a molti corpi. Gli strati successivi sono Fully Connected (FC) e apprendono le mappature tra le caratteristiche estratte e la proprietà fisica di interesse. L'architettura CGCNN è costituita da strati convoluzionali, pooling e FC, mentre l'architettura GeoCGNN si basa su Message Passing Neural Network (MPNN), Gated Convolution e Multilayer Perceptron (MLP). CGCNN e GeoCGNN sono stati implementati utilizzando *Pytorch*, una robusta libreria che permette di eseguire reti neurali (NN) su CPU e GPU di architetture hardware multiple ed eterogenee. Entrambe le reti sono state addestrate con un ampio set di dati archiviato in Red Hat Ceph Storage (bucket S3), una soluzione di storage aperta, scalabile e semplificata progettata per l'analisi dei dati, l'intelligenza artificiale/Machine Learning (AI/ML) e i carichi di lavoro emergenti. Pertanto, è stato necessario modificare entrambi i codici delle reti, disponibili su GitHub, per accedere e gestire i dati.

È stato creato un ambiente virtuale basato su container per impacchettare il codice e tutte le sue dipendenze, in modo che l'applicazione venga eseguita in modo rapido e affidabile da un ambiente di elaborazione all'altro. In particolare, il motore Singularity è stato utilizzato per costruire immagini di container e per eseguire applicazioni complesse su cluster HPC in modo semplice, portatile e riproducibile. La procedura di costruzione è stata completata tramite CRESCOWARE, un gateway basato su container per applicazioni HPC/AI.

In questo studio, esploriamo un ampio spazio combinatorio di materiali candidati per batterie catodiche generando numerose alternative attraverso la sostituzione casuale di elementi cristallini. Per valutare le proprietà di questo ampio set di dati, costituito da sistemi generati drogando la struttura iniziale di NaMnO_2 , abbiamo adottato un approccio data-driven. Le reti neurali a grafo cristallino geometrico (GeoCGNN) sono state utilizzate per prevedere le energie di formazione dei cristalli. GeoCGNN migliora la precedente versione delle reti neurali convoluzionali a grafo di cristallo (CGCNN) integrando informazioni geometriche spaziali dettagliate nel modello. Entrambe le architetture CGCNN e GeoCGNN sfruttano le reti neurali di passaggio dei messaggi (MPNN) e la convoluzione gate. L'input del modello GeoCGNN è un grafo generato sopra la struttura cristallina fornita come Crystallographic Information File (CIF) dove i nodi rappresentano gli atomi nel cristallo e gli spigoli rappresentano le interazioni tra di loro. Gli spigoli sono definiti per atomi separati da meno di 8 Å, con un massimo di 12 vicini considerati per ogni atomo. I nodi sono codificati come vettori di funzionalità one-hot in base alle proprietà atomiche, mentre i vettori di funzionalità del bordo sono i vettori di distanza tra atomi interagenti. Inoltre, il modello incorpora descrittori che codificano i vettori del reticolo e il volume delle celle. L'output è tipicamente una proprietà scalare del materiale. In questo caso, la proprietà target è l'energia di formazione.

La rete è stata addestrata su un ampio set di dati di materiali ottenuti dal database Materials Project (MP) escludendo solo le strutture contenenti gas nobili. I dati sono stati archiviati su Red Hat Ceph Storage, una soluzione su misura per l'analisi dei dati e l'intelligenza artificiale. L'80% del set di dati, composto da 126.162 materiali, è stato utilizzato per la formazione, mentre il restante 20% è stato suddiviso equamente per la convalida e i test. Il processo di addestramento ha utilizzato il metodo Stochastic Gradient Descent (SGD) con un tasso di apprendimento di 0,001, una dimensione del batch di 64 e l'errore quadratico medio (MSE) come funzione di perdita per l'attività di regressione. L'addestramento è stato eseguito per un massimo di 500 epoche, con gli iperparametri più performanti selezionati in base all'errore di convalida. Durante la fase di inferenza, il modello addestrato è stato utilizzato per stimare le energie di formazione del materiale candidato per le batterie catodiche. Un totale di 8.071 strutture cristalline è stato generato casualmente in base alle regole delineate nel report di LA1.20 (D1.23). Poiché queste nuove strutture non sono presenti nel database MP, sono stati eseguiti calcoli DFT per confrontare i valori di energia di formazione ottenuti con quelli previsti dal modello. Per gestire in modo efficiente il gran numero di materiali, sia la generazione casuale delle strutture cristalline che i calcoli DFT sono stati automatizzati utilizzando il flusso di lavoro anche descritto nel report precedente (D1.23).

7.3 Metodo computazionale

Per caratterizzare le proprietà strutturali ed elettroniche dei materiali oggetto di studio, è stato adottato un approccio basato su principi primi utilizzando il codice PWSCF (Plane-Wave Self-Consistent Field), che si basa sulla teoria del funzionale densità (DFT) e fa parte della suite Quantum ESPRESSO. PWSCF conduce vari tipi di calcoli autoconsistenti delle proprietà della struttura elettronica nell'ambito della DFT, utilizzando un set di basi a onde piane (PW) e pseudopotenziali (PP).

L'interazione elettrone-ione è stata descritta utilizzando gli pseudopotenziali ultramorbidi di Rappe-Rabe-Kaxiras-Joannopoulos con correzione del nucleo non lineare, mentre le energie di scambio-correlazione sono state modellate utilizzando l'approssimazione a gradiente generalizzato (GGA) con il funzionale Perdew-Burke-Ernzerhof (PBE). 9, 15, 6, 12 e 10 elettroni sono stati esplicitamente considerati nelle funzioni d'onda per Na, Mn, O, Ti e Ni, rispettivamente. Le funzioni d'onda elettronica sono state ampliate utilizzando una base di onda piana impostata con un cut-off di energia cinetica di 80 Ry, mentre il cutoff di densità di carica è stato impostato a 640 Ry (otto volte maggiore). È stata impiegata una rete k-point Monkhorst-Pack $4 \times 4 \times 4$. Inoltre, è stato applicato uno smearing striatura Marzari-Vanderbilt con una larghezza di 0.01 Ry. Nei nostri calcoli abbiamo considerato la polarizzazione di spin. Sebbene l'approccio GGA (Generalized Gradient Approximation) fornisca previsioni accurate per molte proprietà dei materiali, non è sempre affidabile per tutti i casi. In particolare, i funzionali GGA faticano a descrivere accuratamente le correlazioni elettroniche nei gusci 3d, 4d e 5d parzialmente occupati dei metalli di transizione. Questa limitazione porta spesso alla sottostima della banda proibita di energia negli ossidi dei metalli di transizione. Per affrontare questo problema, è stata adottata la correzione dell'autointerazione U di Hubbard (metodo DFT + U) per considerare la correlazione degli orbitali d. Abbiamo assunto i valori $U_{Mn} = 4.7$ eV, $U_{Ni} = 6.95$ eV e $U_{Ti} = 3.22$ eV per i parametri di Hubbard.

Le condizioni al contorno periodiche (PBC), che consentono la simulazione di un sistema infinitamente esteso, sono state applicate in tutti i calcoli in combinazione con l'approssimazione della supercella. Per determinare valori affidabili per il cutoff dell'energia cinetica, il campionamento del punto k e l'ampiezza di strisciamento, sono stati eseguiti calcoli preliminari di test sugli pseudopotenziali. L'energia totale, le forze e le sollecitazioni sono state verificate per corrispondere a tolleranze di 0.001 Ry/atomo, 0.0001 Ry/Bohr e 0.1 kbar, rispettivamente. Il metodo di minimizzazione del gradiente coniugato (CG) è stato impiegato per la minimizzazione dell'energia, con una soglia di convergenza per l'autoconsistenza fissata a 10^{-6} Ry. L'algoritmo quasi-Newtoniano di Broyden-Fletcher-Goldfarb-Shanno (BFGS) è stato utilizzato per l'ottimizzazione della geometria, che è stata considerata raggiunta quando sono state soddisfatte le seguenti due condizioni: (i) la variazione di energia tra due passi consecutivi di campo self-consistent (SCF) era inferiore a 10^{-4} Ry, e (ii) la forza massima che agisce su un atomo era inferiore a 10^{-3} Ry/Bohr. Abbiamo anche utilizzato l'algoritmo BFGS per il rilassamento a cella variabile, aggiungendo la condizione aggiuntiva che tutti i componenti del tensore di stress siano inferiori a 0.05 GPa.

Per valutare l'impatto del drogaggio nella struttura di partenza (ossido stratificato P2 NaMnO₂), sono state costruite diverse strutture sostituendo gli atomi di Mn (da 1 fino a 12) con atomi di Ti e Ni. Questi sistemi sono stati ispezionati calcolando l'energia di formazione.

L'energia di formazione E_f viene calcolata utilizzando la seguente formula:

$$E_f = \frac{E_{sys}^{tot} - n_{Na}E_{Na}^{bulk} - n_{Mn}E_{Mn}^{bulk} - n_{Ti}E_{Ti}^{bulk} - n_{Ni}E_{Ni}^{bulk} - \frac{1}{2}n_OE_{O_2}}{N_{tot}} \quad (1)$$

Dove E_{sys}^{tot} è l'energia totale del sistema, N_{tot} è il numero totale di atomi contenuti nella supercella, E_X^{bulk} è l'energia totale per atomo della specie X (Na, Mn, Ti, Ni e O), e n_X è il numero di atomi della specie X.

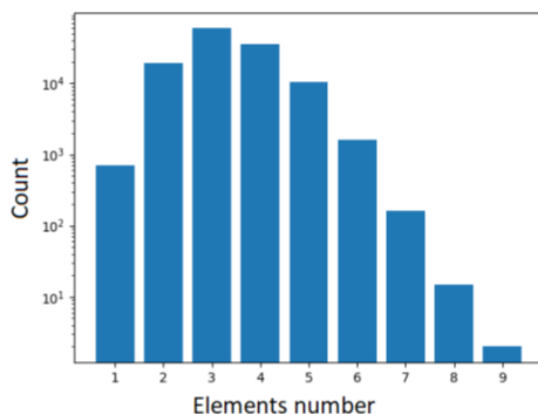
I calcoli sono stati gestiti dal flusso di lavoro AiiDA.

7.4 Dataset

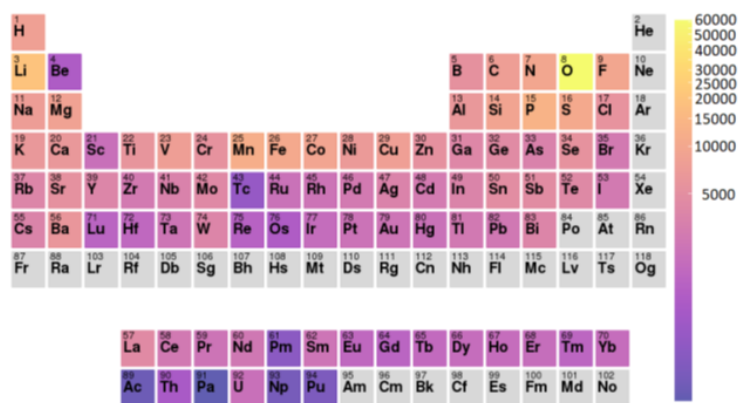
L'apprendimento automatico nella scienza dei materiali si occupa principalmente dell'apprendimento supervisionato. Il successo di tali metodi dipende principalmente dalla quantità e dalla qualità dei dati disponibili. Questa sfida è stata recentemente affrontata da database cristallografici mondiali basati sui principi FAIR, come Materials Project (MP), la Materials genome initiative, NOMAD, Materials Cloud, Inorganic Crystal Structure Database of National Institute of Standards and Technology) contenenti informazioni su numerose proprietà di materiali noti.

Il nostro set di dati è costituito da 126.162 coppie di energia struttura-formazione della versione di febbraio 2022 del database MP. È stato applicato solo un filtro pre-download, che non include strutture con gas nobili. Le NN apprendono le correlazioni fisiche tra gli elementi, facilitandone il riutilizzo per applicazioni future.

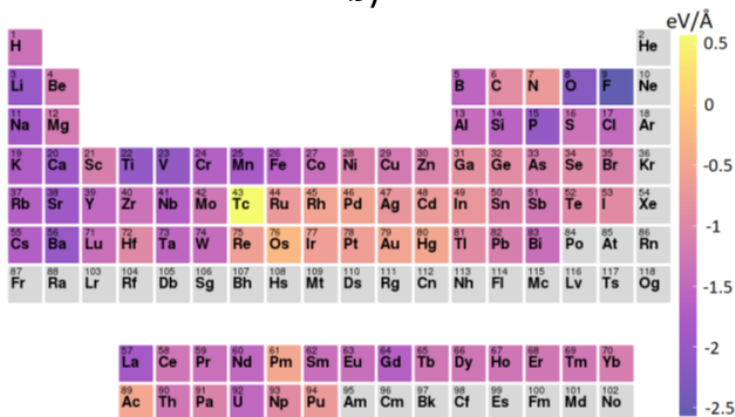
Come mostrato in Figura 1b, l'elemento più presente nel set di dati MP è l'ossigeno (O) con più di 60.000 strutture, mentre gli elementi meno presenti sono l'attinio (Ac) e il protoattinio (Pa) con circa 300 strutture. Inoltre, come illustrato in Figura 1c, l'elemento con la più alta energia media di formazione è il tecnezio (Tc) con più di 0,5 eV per atomo, mentre l'elemento con la più bassa energia media di formazione è il fluoro (F) con meno di -2,5 eV per atomo. Gli elementi di colore grigio non sono presenti nel set di dati MP. I metalli di transizione più interessanti sono il vanadio (V) e il titanio (Ti) con meno di -2.0 eV per atomo.



a)



b)



c)

Figura 1. Statistiche dal set di dati Materials Project (MP). (a) Conteggio degli elementi unici racchiusi nelle strutture su base logaritmica (asse y). (b) Frequenza degli elementi nel dataset su base logaritmica. (c) Energia media di formazione per ciascun elemento. Gli elementi di colore grigio non sono presenti nel set di dati MP.

8 Contributo delle eventuali consulenze alle attività sopra descritte

Le presenti attività sono state effettuate al di fuori dei contratti di consulenza.

9 Pubblicazioni scientifiche

- 1) Study of Cathode Materials for Na-Ion Batteries: Comparison Between Machine Learning Predictions and Density Functional Theory Calculations. Claudio Ronchetti, Sara Marchio, Francesco Buonocore, Simone Giusepponi, Sergio Ferlito, Massimo Celino; Batteries 2024, 10, 431, doi: 10.3390/batteries10120431
- 2) Materials science using machine learning. Claudio Ronchetti, Marco Puccini, Sergio Ferlito, Simone Giusepponi, Filippo Palombi, Francesco Buonocore, Massimo Celino. High Performance Computing on CRESCO Infrastructure: research activities and results 2022. December 2023. ISBN: 978-88-8286-458-3
- 3) Machine learning techniques for data analysis in materials science. Claudio Ronchetti; Marco Puccini; Sergio Ferlito; Simone Giusepponi; Filippo Palombi; Francesco Buonocore, 2022 AEIT International Annual Conference (AEIT), Rome, Italy, 2022, pp. 1-6, doi: 10.23919/AEIT56783.2022.9951839
- 4) La piattaforma IEMAP (Italian Energy Materials Acceleration Platform): Il ruolo dei dati aperti e condivisi. Conferenza GARR 2023 - Selected Papers Volume. Claudio Ronchetti, Sergio Ferlito, Marco Puccini, Simone Giusepponi, Francesco Buonocore, Massimo Celino, Sara Marchio, Giovanni Ponti, doi: 10.26314/GARR-Conf23-proceedings-02

10 Eventi di disseminazione

- 1) Intervento orale "Advances in Na-Ion Battery Cathode Materials: Comparison of DFT and Machine Learning Approaches" nel Simposio "Automation and high throughput research" (speaker: Francesco Buonocore) nella conferenza Nanoinnovation 2024 (Facoltà di Ingegneria Civile e Industriale dell'Università "La Sapienza" di Roma, 12/09/2024)
- 2) Intervento orale "Tecniche AI per la scienza dei materiali" (speaker: Marco Catillo) nel Workshop "Evento Finale Progetto IEMAP" (Roma Eventi "Fontana di Trevi", 03/12/2024)
- 3) Intervento orale "Modellistica molecolare accelerata" (speaker: Francesco Buonocore) nel Workshop "Evento Finale Progetto IEMA" (Roma Eventi "Fontana di Trevi", 03/12/2024)
- 4) Special Issue "Advances in Data-Driven and Learning Methods Applied to Battery Technologies", Guest Editors: Francesco Buonocore, Massimo Celino, Batteries, https://www.mdpi.com/journal/batteries/special_issues/9E38387LYK (May, 2025)
- 5) Intervento orale "Combining Open Data and Computational Workflows to Accelerate Battery Cathode Materials Discovery" nella sessione "Digital Discovery" (speaker: Francesco Buonocore) della European Conference in Computational & Theoretical Chemistry (Napoli, 15-18/09/2025).

11 Descrizione dei risultati ottenuti

11.1 Introduzione

Con l'avanzare della tecnologia globale, l'umanità deve affrontare sfide significative in termini di inquinamento e impatto climatico. L'attuale traiettoria è insostenibile; se da un lato il progresso tecnologico migliora la qualità della nostra vita, dall'altro mette in pericolo il pianeta. È necessario un cambiamento critico, dall'affidarsi ai combustibili fossili come il petrolio e il carbone per la produzione di energia all'adozione di fonti di energia rinnovabile [1]. L'energia rinnovabile consiste principalmente nella conversione dell'energia proveniente dal sole, dal vento e dagli oceani in elettricità. Tuttavia, queste fonti di energia non producono incessantemente energia, creando la necessità di soluzioni efficaci per l'accumulo di energia.

Negli ultimi anni, le batterie agli ioni di litio (LIB) hanno avuto un grande successo nel soddisfare le esigenze di accumulo di energia [2]. Tuttavia, con il crescente fabbisogno energetico, in particolare in settori come i trasporti in cui il passaggio dai combustibili fossili all'elettricità sta accelerando, le LIB devono affrontare due sfide principali: l'aumento del costo del litio a causa della sua disponibilità limitata rispetto alla crescente domanda; problemi di sicurezza legati alla reattività del litio con l'ossigeno. Le batterie agli ioni di sodio (NIB) offrono un'alternativa promettente alle LIB. Il sodio è abbondante e più stabile del litio, il che rende i NIB un'opzione interessante [3]. Inoltre, le NIB sono più sicure dei LIB a causa della minore reattività del sodio rispetto al litio, che mitiga i rischi associati alle batterie infiammabili. In realtà, le NIB sono batterie ricaricabili analoghe alle LIB, in cui entrambi i catodi utilizzano il meccanismo di reazione di intercalazione. I catodi a base di sodio hanno elevati potenziali operativi e capacità elevate, potenzialmente paragonabili a quelli a base di litio [4-6]. Di conseguenza, la combinazione di prestazioni elevate e costi inferiori, insieme alla capacità di sfruttare la tecnologia esistente e i metodi di produzione sviluppati per le LIB senza incorrere in spese aggiuntive, ha alimentato un crescente interesse per questi sistemi.

Il materiale catodico per il sodio a base di ossido lamellare di metallo di transizione (M) (LTMO) Na_xMO_2 è una delle classi di materiali più promettenti da considerare. [7, 8]. Infatti, i materiali Na_xMO_2 presentano transizioni di fase reversibili che migliorano la stabilità delle batterie durante i cicli di carica-scarica. Gli LTMO sono classificati come On e Pn a seconda della fase cristallina ottaedrica e prismatica, rispettivamente.

Diverse strategie sono state proposte per aumentare la densità di energia e per migliorare la stabilità prevenendo le distorsioni di Jahn-Teller [9, 10] nei materiali catodici NIB LTMO, tra cui la modulazione strutturale, la modifica della superficie e il drogaggio elementare [11-15].

La scienza è stata caratterizzata da quattro paradigmi: modelli empirici, formulazione di leggi fisico/chimiche, approcci computazionali basati sulla fisica e tecniche basate sui dati. Questo lavoro si basa sullo sfruttamento del quarto paradigma al fine di accelerare la scoperta scientifica. Il Machine Learning (ML) è diventato una tecnica ampiamente utilizzata, che consente di accelerare il proprio lavoro utilizzando algoritmi addestrati. Nella scienza dei materiali, il ML viene applicato per prevedere le proprietà dei materiali, riducendo significativamente il tempo e le risorse computazionali necessarie rispetto ai tradizionali modelli fisici numerici, che sono spesso ad alta intensità di risorse e si basano sui moderni supercomputer [16].

Esistono molte applicazioni in cui i dati sono generati da domini non euclidei, rappresentati come grafi con relazioni complesse e interdipendenze tra oggetti come i social network [17], i recommender systems [18] e le scienze della vita [19]. Le reti neurali a grafo (GNN) sono una classe di metodi di deep learning progettati per eseguire inferenze sui dati descritti dai grafi. Le GNN forniscono un modo semplice per eseguire attività di regressione e classificazione a livello di nodo e di bordo.

Negli ultimi anni, ci sono stati rapidi progressi nello sviluppo di reti neurali basate sul deep learning al fine di prevedere le proprietà dei materiali. Tali reti hanno prodotto previsioni accurate e rapide. La molecola o materiale cristallino è rappresentato come un grafo in cui gli atomi costituiscono i nodi e i loro legami atomistici corrispondono al bordo. Nella scienza dei materiali sono state sviluppate reti neurali basate su architetture GNN come SchNet [20-21], Crystal Graph Convolutional Neural Networks (CGCNN) [22],

MatERials Graph Network (MEGNet) [23], Crystal Graph Convolutional Neural Networks (iCGCNN) [24], OrbNet [25], Atomistic Line Graph Neural Network (ALIGNN) [26] e Hydra GNN [27].

Nel presente rapporto proponiamo uno studio computazionale per migliorare la stabilità dei materiali catodici Na_xMO_2 mediante drogaggio sostitutivo di elementi Mn con due diversi metodi: 1) calcoli di principi primi high-throughput basati sulla teoria del funzionale densità (DFT); 2) tecniche di machine learning basate sull'addestramento di una rete neurale convoluzionale (CNN) a partire da un dataset generato da calcoli DFT. La stabilità è determinata con calcoli DFT o previsioni CNN per individuare l'energia di formazione (i cristalli più stabili hanno un'energia di formazione inferiore). Abbiamo confrontato i due metodi per valutare la capacità predittiva della CNN durante l'addestramento. Presentiamo anche una analisi predittiva del potenziale redox dei materiali catodici, proprietà utile per stimare l'efficienza energetica della batteria. La struttura di partenza è stata determinata sulla base di studi preliminari. Questa struttura è NaMnO_2 e i calcoli vengono eseguiti in una supercella composta da 48 atomi. Il processo consiste nel drogaggio della struttura di partenza sostituendo il Manganese con Titanio e Nichel, ed il calcolo dell'energia di formazione e del potenziale redox per ciascuna struttura. I risultati vengono visualizzati su grafici ternari per evidenziare le aree ottimali. In questa attività sono state utilizzate due reti, CGCNN e la più recente Geometric-Information-Enhanced Crystal Graph Network (GeoCGNN) [28], per prevedere le proprietà del materiale.

Nel prossimo paragrafo 11.2 saranno illustrate le metodologie adottate, questa volta con i riferimenti di letteratura, e quindi saranno delineate le architetture delle reti neurali basate sulla teoria dei grafi, l'approccio computazionale ab initio e il dataset da cui si è partiti per addestrare le reti. Nei paragrafi successivi saranno riportati e discussi: nel paragrafo 11.3 i risultati numerici dell'addestramento; nel paragrafo 11.4 i risultati numerici dell'inferenza; nel paragrafo 11.5 il confronto tra le predizioni e i calcoli ab initio; nel paragrafo 11.6 le diverse nature degli algoritmi predittivi addestrabili partendo da un dataset di misure sperimentali. Infine, nel paragrafo 11.7 trarremo le conclusioni.

11.2 Descrizione dei metodi usati

11.2.1 Architetture informatiche

Nella presente attività si è lavorato su due architetture di rete neurale, CGCNN e GeoCGNN. Queste reti sono state scelte perché, in primo luogo, la CGCNN è un punto di riferimento per la scienza dei materiali e si è consolidata negli ultimi anni, e in secondo luogo era necessaria una versione potenziata che non richiedesse tempi di gestione aggiuntivi (GeoCGNN). Entrambi sono basati su architetture GNN. Le reti neurali e le differenze tra loro sono correlate alla rappresentazione del grafo cristallino e agli strati (layers) della rete neurale.

Rappresentazione del grafo cristallino. Molte ricerche di chimica informatica hanno dimostrato l'importanza critica dei descrittori molecolari, che sono noti per influenzare l'accuratezza del modello più della scelta dell'algoritmo di apprendimento automatico [29-30]. Per CGCNN, ogni nodo i del grafo cristallino G è caratterizzato da nove vettori di caratteristiche v_i , ciascuno per una proprietà atomica in cui le proprietà continue sono suddivise in dieci categorie, mentre le proprietà discrete possono ovviamente essere codificate automaticamente. Allo stesso modo, ogni spigolo $(i, j)_k$ è rappresentato da un vettore di caratteristiche $u_{(i,j)_k}$, corrispondente all'interazione del legame k -esimo tra i nodi i e j , ed è costruito in modo analogo al descrittore Property-Labeled Material Fragments (PLMF), un metodo ispirato da Ref. [31-32]. I PMLF si differenziano per le proprietà di riferimento locali, che includono proprietà generali, misurate e derivate. In primo luogo, la struttura cristallina viene analizzata alla ricerca di vicini atomici tramite tassellatura di Voronoi e quindi il grafo periodico risultante viene scomposto in semplici sottografi. Ci sono diversi sottografi da considerare in generale, ma sono stati limitati a un massimo di tre archi consecutivi e non ripetitivi nel sottografo al fine di frenare la complessità del vettore descrittore finale. Analogamente a CGCNN, GeoCGNN definisce i vicini di ogni atomo all'interno di un raggio specifico risultando in un grafo periodico e definisce le rappresentazioni

dei nodi e degli spigoli sul grafo. I nodi sono rappresentati come vettore di caratteristiche one-hot che dipende dal numero atomico, mentre gli archi sono rappresentati come vettore di distanza tra gli atomi (grafo diretto). Inoltre, GeoCGNN utilizza un descrittore che registra il vettore reticolare e il volume della cella per mantenere le informazioni globali.

Rete neurale. CGCNN e GeoCGNN immettono direttamente la struttura atomica e la convertono in un grafo, dove gli atomi sono interpretati come nodi e i legami interatomici sono interpretati come spigoli. Principalmente, entrambe le reti si basano su due insiemi di strati (layers): il primo insieme di strati apprende le caratteristiche comuni a tutte le proprietà del materiale e l'ultimo insieme di strati è separato in più capi per apprendere le caratteristiche specifiche di ciascuna proprietà del materiale. Gli strati convoluzionali del grafo vengono utilizzati per estrarre caratteristiche comuni rilevanti da interazioni di vicini a coppie e, attraverso più strati, rappresentano anche interazioni a molti corpi. Gli strati successivi sono Fully Connected (FC) e apprendono le mappature tra le caratteristiche estratte e la proprietà fisica di interesse. L'architettura CGCNN è costituita da strati convoluzionali, pooling e FC, mentre l'architettura GeoCGNN si basa su Message Passing Neural Network (MPNN), Gated Convolution e Multilayer Perceptron (MLP). CGCNN e GeoCGNN sono stati implementati utilizzando *Pytorch* [33-34], una robusta libreria che permette di eseguire reti neurali (NN) su CPU e GPU di architetture hardware multiple ed eterogenee. Entrambe le reti sono state addestrate con un ampio set di dati archiviato in Red Hat Ceph Storage (bucket S3) [35], una soluzione di storage aperta, scalabile e semplificata progettata per l'analisi dei dati, l'intelligenza artificiale/Machine Learning (AI/ML) e i carichi di lavoro emergenti. Pertanto, è stato necessario modificare entrambi i codici delle reti, disponibili su GitHub [36-37], per accedere e gestire i dati.

È stato creato un ambiente virtuale basato su container per impacchettare il codice e tutte le sue dipendenze, in modo che l'applicazione venga eseguita in modo rapido e affidabile da un ambiente di elaborazione all'altro. In particolare, il motore Singularity [38] è stato utilizzato per costruire immagini di container e per eseguire applicazioni complesse su cluster HPC in modo semplice, portatile e riproducibile. La procedura di costruzione è stata completata tramite CRESCOWARE [39], un gateway basato su container per applicazioni HPC/AI.

In questo studio, esploriamo un ampio spazio combinatorio di materiali candidati per batterie catodiche generando numerose alternative attraverso la sostituzione casuale di elementi cristallini. Per valutare le proprietà di questo ampio set di dati, costituito da sistemi generati drogando la struttura iniziale di NaMnO_2 , abbiamo adottato un approccio data-driven. Le reti neurali a grafo cristallino geometrico (GeoCGNN) [28] sono state utilizzate per prevedere le energie di formazione dei cristalli. GeoCGNN migliora la precedente versione delle reti neurali convoluzionali a grafo di cristallo (CGCNN) [22] integrando informazioni geometriche spaziali dettagliate nel modello. Entrambe le architetture CGCNN e GeoCGNN sfruttano le reti neurali di passaggio dei messaggi (MPNN) [40] e la convoluzione gate. L'input del modello GeoCGNN è un grafo generato sopra la struttura cristallina fornita come Crystallographic Information File (CIF) dove i nodi rappresentano gli atomi nel cristallo e gli spigoli rappresentano le interazioni tra di loro. Gli spigoli sono definiti per atomi separati da meno di 8 Å, con un massimo di 12 vicini considerati per ogni atomo. I nodi sono codificati come vettori di funzionalità one-hot in base alle proprietà atomiche, mentre i vettori di funzionalità del bordo sono i vettori di distanza tra atomi interagenti. Inoltre, il modello incorpora descrittori che codificano i vettori del reticolo e il volume delle celle. L'output è tipicamente una proprietà scalare del materiale. In questo caso, la proprietà target è l'energia di formazione.

La rete è stata addestrata su un ampio set di dati di materiali ottenuti dal database Materials Project (MP) [41] escludendo solo le strutture contenenti gas nobili. I dati sono stati archiviati su Red Hat Ceph Storage [35], una soluzione su misura per l'analisi dei dati e l'intelligenza artificiale. L'80% del set di dati, composto da 126.162 materiali, è stato utilizzato per la formazione, mentre il restante 20% è stato suddiviso equamente

per la convalida e i test. Il processo di addestramento ha utilizzato il metodo Stochastic Gradient Descent (SGD) [42] con un tasso di apprendimento di 0,001, una dimensione del batch di 64 e l'errore quadratico medio (MSE) come funzione di perdita per l'attività di regressione. L'addestramento è stato eseguito per un massimo di 500 epoche, con gli iperparametri più performanti selezionati in base all'errore di convalida. Durante la fase di inferenza, il modello addestrato è stato utilizzato per stimare le energie di formazione del materiale candidato per le batterie catodiche. Un totale di 8.071 strutture cristalline è stato generato casualmente in base alle regole delineate nel report di LA1.20 (D1.23). Poiché queste nuove strutture non sono presenti nel database MP, sono stati eseguiti calcoli DFT per confrontare i valori di energia di formazione ottenuti con quelli previsti dal modello. Per gestire in modo efficiente il gran numero di materiali, sia la generazione casuale delle strutture cristalline che i calcoli DFT sono stati automatizzati utilizzando il flusso di lavoro anche descritto nel report precedente (D1.23).

11.2.2 Metodo computazionale

Per caratterizzare le proprietà strutturali ed elettroniche dei materiali oggetto di studio, è stato adottato un approccio basato su principi primi utilizzando il codice PWSCF (Plane-Wave Self-Consistent Field), che si basa sulla teoria del funzionale densità (DFT) e fa parte della suite Quantum ESPRESSO [43-44]. PWSCF conduce vari tipi di calcoli autoconsistenti delle proprietà della struttura elettronica nell'ambito della DFT [45], utilizzando un set di basi a onde piane (PW) e pseudopotenziali (PP).

L'interazione elettrone-elettrone è stata descritta utilizzando gli pseudopotenziali ultramorbidi di Rappe-Rabe-Kaxiras-Joannopoulos con correzione del nucleo non lineare, mentre le energie di scambio-correlazione sono state modellate utilizzando l'approssimazione a gradiente generalizzato (GGA) con il funzionale Perdew-Burke-Ernzerhof (PBE) [46]. 9, 15, 6, 12 e 10 elettroni sono stati esplicitamente considerati nelle funzioni d'onda per Na, Mn, O, Ti e Ni, rispettivamente. Le funzioni d'onda elettronica sono state ampliate utilizzando una base di onda piana impostata con un cut-off di energia cinetica di 80 Ry, mentre il cutoff di densità di carica è stato impostato a 640 Ry (otto volte maggiore). È stata impiegata una rete k-point Monkhorst-Pack $4 \times 4 \times 4$ [47]. Inoltre, è stato applicato uno smearing striatura Marzari-Vanderbilt [48] con una larghezza di 0.01 Ry. Nei nostri calcoli abbiamo considerato la polarizzazione di spin. Sebbene l'approccio GGA (Generalized Gradient Approximation) fornisca previsioni accurate per molte proprietà dei materiali, non è sempre affidabile per tutti i casi. In particolare, i funzionali GGA faticano a descrivere accuratamente le correlazioni elettroniche nei gusci 3d, 4d e 5d parzialmente occupati dei metalli di transizione. Questa limitazione porta spesso alla sottostima della banda proibita di energia negli ossidi dei metalli di transizione. Per affrontare questo problema, è stata adottata la correzione dell'autointerazione U di Hubbard (metodo DFT + U) per considerare la correlazione degli orbitali d [49]. Abbiamo assunto i valori $U_{Mn} = 4.7$ eV, $U_{Ni} = 6.95$ eV e $U_{Ti} = 3.22$ eV per i parametri di Hubbard.

Le condizioni al contorno periodiche (PBC), che consentono la simulazione di un sistema infinitamente esteso, sono state applicate in tutti i calcoli in combinazione con l'approssimazione della supercella. Per determinare valori affidabili per il cutoff dell'energia cinetica, il campionamento del punto k e l'ampiezza di strisciamento, sono stati eseguiti calcoli preliminari di test sugli pseudopotenziali. L'energia totale, le forze e le sollecitazioni sono state verificate per corrispondere a tolleranze di 0.001 Ry/atomo, 0.0001 Ry/Bohr e 0.1 kbar, rispettivamente. Il metodo di minimizzazione del gradiente coniugato (CG) è stato impiegato per la minimizzazione dell'energia, con una soglia di convergenza per l'autoconsistenza fissata a 10^{-6} Ry. L'algoritmo quasi-Newtoniano di Broyden-Fletcher-Goldfarb-Shanno (BFGS) è stato utilizzato per l'ottimizzazione della geometria, che è stata considerata raggiunta quando sono state soddisfatte le seguenti due condizioni: (i) la variazione di energia tra due passi consecutivi di campo self-consistent (SCF) era inferiore a 10^{-4} Ry, e (ii) la forza massima che agisce su un atomo era inferiore a 10^{-3} Ry/Bohr. Abbiamo anche utilizzato l'algoritmo BFGS per il rilassamento a cella variabile, aggiungendo la condizione aggiuntiva che tutti i componenti del tensore di stress siano inferiori a 0.05 GPa.

Per valutare l'impatto del drogaggio nella struttura di partenza (ossido stratificato $P2\ NaMnO_2$), sono state costruite diverse strutture sostituendo gli atomi di Mn (da 1 fino a 12) con atomi di Ti e Ni. Questi sistemi sono stati ispezionati calcolando l'energia di formazione.

L'energia di formazione E_f viene calcolata utilizzando la formula (1). I calcoli sono stati gestiti dal flusso di lavoro AiIDA [50-51]

11.2.3 Dataset

L'apprendimento automatico nella scienza dei materiali si occupa principalmente dell'apprendimento supervisionato. Il successo di tali metodi dipende principalmente dalla quantità e dalla qualità dei dati disponibili. Questa sfida è stata recentemente affrontata da database cristallografici mondiali basati sui principi FAIR [52], come Materials Project (MP) [41], la Materials genome initiative [53], NOMAD [54], Materials Cloud [55], Inorganic Crystal Structure Database of National Institute of Standards and Technology [56]) contenenti informazioni su numerose proprietà di materiali noti.

Il nostro set di dati è costituito da 126.162 coppie di energia struttura-formazione della versione di febbraio 2022 del database MP. È stato applicato solo un filtro pre-download, che non include strutture con gas nobili.

Le NN apprendono le correlazioni fisiche tra gli elementi, facilitandone il riutilizzo per applicazioni future.

Come mostrato in Figura 1b, l'elemento più presente nel set di dati MP è l'ossigeno (O) con più di 60.000 strutture, mentre gli elementi meno presenti sono l'attinio (Ac) e il protoattinio (Pa) con circa 300 strutture. Inoltre, come illustrato in Figura 1c, l'elemento con la più alta energia media di formazione è il tecnezio (Tc) con più di 0,5 eV per atomo, mentre l'elemento con la più bassa energia media di formazione è il fluoro (F) con meno di -2,5 eV per atomo. Gli elementi di colore grigio non sono presenti nel set di dati MP. I metalli di transizione più interessanti sono il vanadio (V) e il titanio (Ti) con meno di -2.0 eV per atomo.

11.3 Risultati numerici addestramento

Questa sezione presenta i risultati numerici di previsione dell'energia di formazione per i materiali allo stato solido. In primo luogo, sono state confrontate le prestazioni predittive di diverse configurazioni di reti CGCNN e GeoCGNN. In secondo luogo, l'inferenza è stata eseguita utilizzando il miglior modello sul set di dati dello spazio combinatorio al fine di rilevare l'area ottimale.

I modelli addestrati sono costituiti da cinque CGCNN che utilizzano 1, 2, 3, 5 e 10 strati convoluzionali e un GeoCGNN (modello predefinito). Un raggio di taglio di 8 Å e un massimo di 12 atomi sono utilizzati per definire i vicini locali.

I modelli sono stati addestrati utilizzando il metodo Stochastic Gradient Descent (SGD) [42] con un tasso di apprendimento pari a 0.001, dimensioni dei batch di 64, errore quadratico medio (MSE) come funzione di perdita per l'attività di regressione e un numero massimo di epoche impostato su 500. L'arresto anticipato viene eseguito per interrompere l'addestramento quando la funzione di perdita di convalida non diminuisce per diverse epoche consecutive, poiché questo è un sintomo che mostra che è molto improbabile che ulteriori epoche riducano il valore della funzione di perdita.

I modelli sono stati addestrati con il set di dati MP descritto nella sezione precedente per prevedere l'energia di formazione, che viene riscalata per la standardizzazione, eq. (2). La standardizzazione Z è un metodo di scala in cui i valori sono centrati attorno alla media $\bar{x}$ con una deviazione standard unitaria σ .

$$Z(x) = \frac{x - \bar{x}}{\sigma} \quad (2)$$

Il set di dati è diviso in tre parti: addestramento, convalida e test set, utilizzando la suddivisione 80, 10 e 10% rispettivamente.

Ogni modello è stato addestrato sul set di addestramento per 500 epoche e sono stati selezionati la configurazione e gli iperparametri con l'errore di convalida più basso. Infine, viene calcolato l'errore assoluto medio (MAE) del test. I modelli sono stati addestrati su un nodo dotato di una GPU Nvidia Quadro RTX 6000 disponibile nell'infrastruttura HPC CRESCO/ENEAGRID [57, 58].

La Tabella 1 confronta l'accuratezza predittiva di diverse configurazioni utilizzate per prevedere l'energia di formazione. La tabella mostra il MAE medio e il coefficiente di determinazione (R^2) sul set di test. Nel complesso, il modello GeoCGNN supera le configurazioni CGCNN poiché il MAE di GeoCGNN è almeno la metà di quello di altri modelli. Inoltre, l'aumento della complessità di GeoCGNN si traduce in tempi di addestramento più lunghi.

Tabella 1. MAE delle proprietà energetiche di formazione.

Modello (a)	MAE (eV atom ⁻¹)	R^2	Tempo di training
CGCNN (1)	0.145	0.951	11h24m
CGCNN (2)	0.117	0.965	11h32m
CGCNN (3)	0.106	0.966	11h44m
CGCNN (5)	0.095	0.970	13h32m
CGCNN (10)	0.089	0.971	21h33m
GeoCGNN	0.044	0.989	29h05m

^aNumero di strati (layers) convoluzionali.

11.4 Risultati numerici dell'inferenza

Una volta ottenuto il modello migliore, è stato utilizzato per eseguire l'inferenza per tutte le strutture trovate nello spazio combinatorio di interesse. In primo luogo, lo spazio combinatorio è stato definito utilizzando una funzione ricorsiva a partire dalla struttura NaMnO₂ con 48 atomi. L'algoritmo 1 rappresenta lo pseudocodice della funzione ricorsiva in cui i suoi input sono una molecola $mol_{(t)}$ e un elenco di posizioni Mn $pos_{(t)}$ al tempo t . Se l'elenco delle posizioni $pos_{(t)}$ è vuoto, allora la molecola di input $mol_{(t)}$ è una struttura drogata e la funzione la restituisce. In caso contrario, si chiama per ogni TM_x del metallo di transizione in cui gli input sono:

- Molecola $mol_{(t+1)}$: è una copia della molecola $mol_{(t)}$ con un metallo di transizione TM_x posto nel primo elemento di $pos_{(t)}$;
- Elenco degli indici $pos_{(t+1)}$: è $pos_{(t)}$ senza il primo elemento.

I metalli di transizione TM sono un elenco di elementi con manganese, nichel e titanio. L'algoritmo può essere riutilizzato utilizzando altri metalli di transizione.

Algorithm 1 The algorithm of *RECURSIVE* function to define the doped structures

Input: $mol_{(t)}$ is the molecule and $pos_{(it)}$ is the list of Mn positions at the t time

Output: $struct_{doped}$: List of doped structures

```

RECURSIVE ( $mol_{(t)}$ ,  $pos_{(t)}$ )
  List[]  $struct_{doped} \leftarrow$  new List []
  if  $pos_{(t)} \equiv$  List[] then
     $struct_{doped}.insert(mol_{(t)})$ 
  else
    for  $i \leftarrow 1$  to  $length(TM)$  do
       $mol_{(t+1)} \leftarrow mol_{(t)}$ 
       $mol_{(t+1)}[pos_{(t)}.next] = TM_x$ 
       $pos_{(t+1)} \leftarrow pos_{(t)}.drop$ 
       $struct_{doped}.insert(RECURSIVE(mol_{(t+1)},$ 
         $pos_{(t+1)}))$ 
    end for
  end if
  return  $struct_{doped}$ 

```

L'equazione (3) mostra il numero totale di strutture drogate C_{doped} a partire dalla struttura $NaMnO_2$, dove n_{tm} è il numero di metalli di transizione e N_{Mn} è il conteggio degli atomi di Mn. In questo lavoro $n_{tm} = 2$ (nichel e titanio) e $N_{Mn} = 12$. C_{doped} include anche le strutture simmetricamente equivalenti.

$$C_{doped} = (n_{tm} + 1)^{N_{Mn}} = 3^{12} = 531441 \quad (3)$$

In secondo luogo, è stato utilizzato il miglior modello per prevedere l'energia di formazione per ogni struttura drogata. Il tempo di esecuzione per eseguire l'inferenza ha richiesto circa 60 ms per ogni struttura utilizzando 48 CPU su un nodo, che non è lontanamente paragonabile alle 3 ore in media necessarie per eseguire i calcoli DFT utilizzando 576 CPU distribuite su 12 nodi.

Il potenziale redox V , la tensione media tra due limiti di intercalazione (x_2 e x_1) del materiale, è calcolato come segue:

$$V(x_1, x_2) = \frac{E_t(Na_{x_2}MnO_2) - E_t(Na_{x_1}MnO_2) - (x_2 - x_1)E_{Na}}{(x_2 - x_1)F} \quad (4)$$

dove $E_t(Na_xMnO_2)$ è l'energia totale per unità di formula del materiale ospite con intercalazione x , E_{Na} è l'energia totale per unità di formula di Na nella struttura cubica centrata sul corpo e F è la costante di Faraday, quindi il potenziale è espresso in Volt.

L'energia totale E_t è calcolata come:

$$E_t = n_t E_f + \sum_x n_x E_x \quad (5)$$

dove n_x e n_t sono rispettivamente il numero di atomi della specie x e il numero totale di atomi contenuti nella supercella, ed E_x è l'energia totale per atomo della specie x . Entrambe le proprietà dei materiali di interesse (energia di formazione e potenziale redox) sono calcolate e illustrate in Figura 2a e Figura 2b utilizzando il grafico ternario, uno strumento per l'analisi dei dati composizionali nel caso tridimensionale. La libreria Python chiamata *python-ternary* è stata utilizzata per il rendering di grafici ternari [59].

In questo lavoro, i tre vertici dei grafici ternari sono Manganese, Nichel e Titanio e ogni punto grafico è una composizione molecolare del ternario (Mn, Ni, Ti). Poiché ci sono migliaia di strutture per ogni composizione, per ciascuna di esse viene presa la media delle proprietà dei materiali previste. Un'alta percentuale di Titanio determina una minore energia di formazione, mostrata in Figura 2a, mentre un'alta percentuale di Nichel risulta un potenziale redox più elevato, mostrato in Figura 2b. Infine, il problema dell'ottimizzazione multi-obiettivo è formulato come:

$$\max_{x \in X} (MO(x)) = \max_{x \in X} (-E_f(x), V(x)) \quad (6)$$

dove $MO(x)$ è il multi-obiettivo, $E_f(x)$ è l'energia di formazione, $V(x)$ è il potenziale redox e $x \in X$, cioè l'insieme delle composizioni. Poiché minore è l'energia di formazione, maggiore è il materiale ottimo, è stato applicato il segno negativo. Inoltre, entrambe le proprietà del materiale vengono scalate utilizzando la normalizzazione min-max, pesate e sommate per ogni composizione x per calcolare il $MO(x)$ multi-obiettivo come:

$$MO(x) = w_1(-E_f(x)_{scaled}) + w_2V(x)_{scaled} \quad (7)$$

Poiché non esiste una proprietà del materiale preferita, i pesi w_1 e w_2 sono stati posti uguali a 1. L'area ottimale, la parte gialla in Figura 2c, è intorno alla composizione molecolare con atomi di Manganese, Nichel e Titanio pari rispettivamente agli indici 0, 5 e 7. Il codice utilizzato per questa attività è disponibile all'indirizzo <https://gitlab.brindisi.enea.it/claudio.ronchetti/ai4mat>, e può essere usato anche per essere integrato in altri codici.

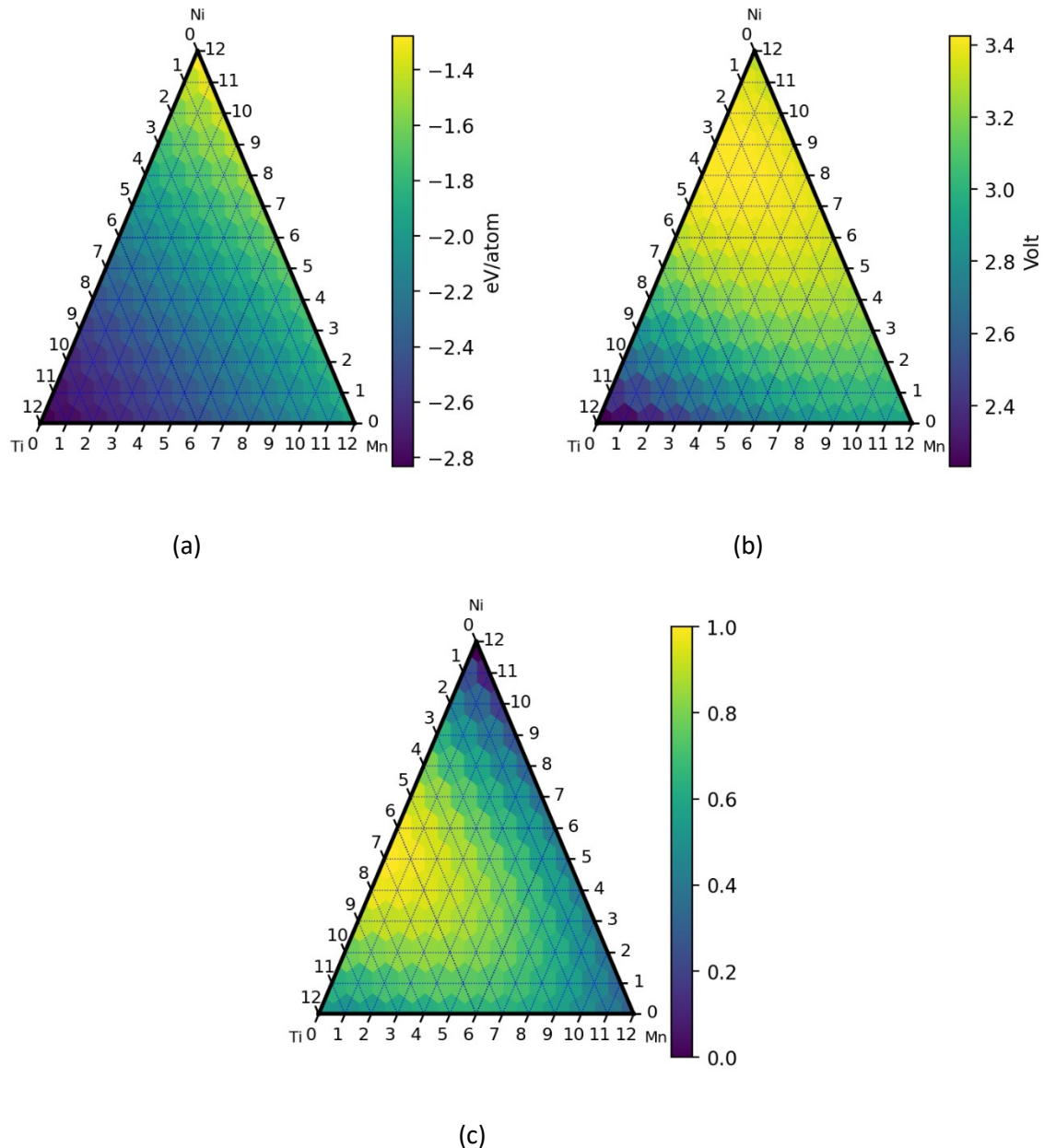


Figura 2. Grafici ternari delle proprietà dei materiali: (a) energia di formazione e (b) potenziale redox, previsto dal modello NN; c) soluzione multiobiettivo. Sugli spigoli è riportato il numero di atomi. Scala cromatica: in (a) la concentrazione ottimale è l'area blu, viceversa in (b) e (c) è l'area gialla.

11.5 Confronto con calcoli ab initio

L'energia di formazione è stata calcolata utilizzando l'equazione (1), per il cristallo base NaMnO_2 , E_f è -2.01 eV per il sistema non rilassato e -2.12 eV per quello rilassato. Per fare un confronto tra le energie di formazione previste e quelle effettive è stato necessario creare un dataset di test utile a questo caso d'uso. La creazione di un buon set di dati di test è fondamentale per valutare le prestazioni e la generalizzabilità di un modello di Deep Learning (DL).

In primo luogo, il set di dati di test dovrebbe rappresentare l'intera distribuzione dei dati che il modello DL incontrerà nell'uso comune [60-61]. Se il set di dati di test non rappresenta la diversità dell'uso reale, le metriche di valutazione saranno fuorvianti. Le regole del flusso di lavoro AiiDA [50-51] hanno reso possibile rappresentare la diversità creando un set di dati bilanciato. In secondo luogo, il set di dati del test dovrebbe essere sufficientemente grande da fornire stime affidabili delle prestazioni [60, 62]. Questo era difficile da ottenere a causa dei tempi elevati richiesti per le simulazioni. Nonostante queste difficoltà, il set di dati contiene più di 500 calcoli. Infine, assicurarsi che il set di dati di test non includa informazioni dal set di dati di addestramento [60, 61]. In questo caso, nessuno dei nostri sistemi calcolati viene memorizzato nel database dei progetti dei materiali.

Il set di dati di test è stato creato tenendo conto di tutti questi principi utilizzando il flusso di lavoro AiiDA. Nella Figura 3, confrontiamo e convalidiamo le previsioni del nostro metodo ML con i risultati dei calcoli DFT. A sinistra, le previsioni di E_f per le 8.071 strutture cristalline non rilassate generate casualmente sono rappresentate in cerchi grigi, mostrando un intervallo negativo di valori compreso tra -3 eV e -1 eV. Inoltre, alcuni sistemi sono evidenziati da cerchi di colori diversi ed etichettati con la loro composizione specifica. Ciò evidenzia che il drogaggio del titanio aumenta la stabilità del sistema (E_f più negativo), mentre il drogaggio del nichel la riduce (E_f meno negativo). Infatti, $\text{Na}_{12}\text{Ti}_{12}\text{O}_{24}$ (cerchio viola) è il sistema con la E_f più bassa (-2.78 eV); al contrario, $\text{Na}_{12}\text{Ni}_{12}\text{O}_{24}$ è il sistema (cerchio verde scuro) con il valore più alto di E_f (-1.20 eV). Il cristallo di base $\text{Na}_{12}\text{Mn}_{12}\text{O}_{24}$ (cerchio nero), che si colloca al centro dell'intervallo (-1.88 eV), è uno stato intermedio di stabilità. La sostituzione di un atomo di manganese con un atomo di titanio ($\text{Na}_{12}\text{Mn}_{11}\text{TiO}_{24}$, cerchio rosso, $E_f = -1.97$ eV) diminuisce l'energia di formazione di 0,09 eV, aumentando così la stabilità del sistema.; al contrario, la sostituzione di un atomo di manganese con un atomo di nichel ($\text{Na}_{12}\text{Mn}_{11}\text{NiO}_{24}$, cerchio marrone, $E_f = -1.84$ eV) aumenta l'energia di formazione di 0,04 eV, rendendo la struttura meno stabile. In Figura 3 sono riportate anche le previsioni per i sistemi $\text{Na}_{12}\text{Mn}_4\text{Ni}_4\text{Ti}_4\text{O}_{24}$ (cerchi arancioni) con uguale composizione dei tre metalli di transizione (Mn, Ti e Ni) e per i sistemi in cui due metalli di transizione hanno la stessa concentrazione, ovvero $\text{Na}_{12}\text{Mn}_3\text{Ni}_6\text{Ti}_3\text{O}_{24}$ (cerchi magenta), $\text{Na}_{12}\text{Mn}_6\text{Ni}_3\text{Ti}_3\text{O}_{24}$ (cerchi verdi) e $\text{Na}_{12}\text{Mn}_3\text{Ni}_3\text{Ti}_6\text{O}_{24}$ (cerchi azzurri). Poiché per queste composizioni esistono diverse configurazioni, oltre all'intervallo di variabilità dell'energia di formazione, sono stati calcolati il valore medio $\langle E_f \rangle$ e le deviazioni standard σ_{E_f} per valutare l'incertezza statistica. Per sistemi $\text{Na}_{12}\text{Mn}_4\text{Ni}_4\text{Ti}_4\text{O}_{24}$: range da -2.10 a -2.00 eV; $\langle E_f \rangle = -2.06$ eV; $\sigma_{E_f} = 0,01$ eV. Per sistemi $\text{Na}_{12}\text{Mn}_3\text{Ni}_6\text{Ti}_3\text{O}_{24}$: range da -1.87 a -1.80 eV; $\langle E_f \rangle = -1.85$ eV; $\sigma_{E_f} = 0.01$ eV. Per sistemi $\text{Na}_{12}\text{Mn}_6\text{Ni}_3\text{Ti}_3\text{O}_{24}$: range da -2.05 a -1.98 eV; $\langle E_f \rangle = -2.02$ eV; $\sigma_{E_f} = 0.02$ eV. Per sistemi $\text{Na}_{12}\text{Mn}_3\text{Ni}_3\text{Ti}_6\text{O}_{24}$: range da -2.30 a -2.24 eV; $\langle E_f \rangle = -2.27$ eV; $\sigma_{E_f} = 0.01$ eV. Nella Figura 3, l'incertezza statistica è indicata da linee continue nere centrate nei valori medi e limitate dalle deviazioni standard. Sul lato destro della Figura 3, i valori calcolati dell'energia di formazione di sistemi non rilassati utilizzando il metodo dei principi primi sono mostrati da simboli a croce colorati. Almeno uno dei sistemi evidenziati a sinistra nella figura è stato analizzato mediante calcoli DFT valutandone l'energia di formazione. In particolare, vediamo che si trova lo stesso ordine di stabilità dell'energia di formazione prevista. Il sistema $\text{Na}_{12}\text{Ti}_{12}\text{O}_{24}$ (croce viola) è il più stabile con $E_f = -2.52$ eV; il sistema $\text{Na}_{12}\text{Ni}_{12}\text{O}_{24}$ (croce verde scuro) è il meno stabile con $E_f = -1,15$ eV. Il cristallo di base $\text{Na}_{12}\text{Mn}_{12}\text{O}_{24}$ (croce nera) ha un valore di energia di formazione pari a -2,01 eV; valore tra quelli corrispondenti al sistema $\text{Na}_{12}\text{Mn}_{11}\text{TiO}_{24}$ (croce rossa; $E_f = -2.03$ eV) e al sistema $\text{Na}_{12}\text{Mn}_{11}\text{NiO}_{24}$ (croce marrone; $E_f = -1.93$ eV), rispettivamente. I valori calcolati dell'energia di formazione dei sistemi $\text{Na}_{12}\text{Mn}_4\text{Ni}_4\text{Ti}_4\text{O}_{24}$ sono (croci arancioni, range da -1.98 a -1.94 eV). Infine, elenchiamo anche i valori calcolati di E_f per i sistemi in cui due metalli di transizione hanno la stessa concentrazione, ovvero $\text{Na}_{12}\text{Mn}_3\text{Ni}_6\text{Ti}_3\text{O}_{24}$ (croci magenta; range da -1.80 a -1.77 eV), $\text{Na}_{12}\text{Mn}_6\text{Ni}_3\text{Ti}_3\text{O}_{24}$ (croci verdi; range da -2.07 a -1.96 eV) e $\text{Na}_{12}\text{Mn}_4\text{Ni}_4\text{Ti}_4\text{O}_{24}$ (croci azzurre; range da -2.15 a -2.12 eV).

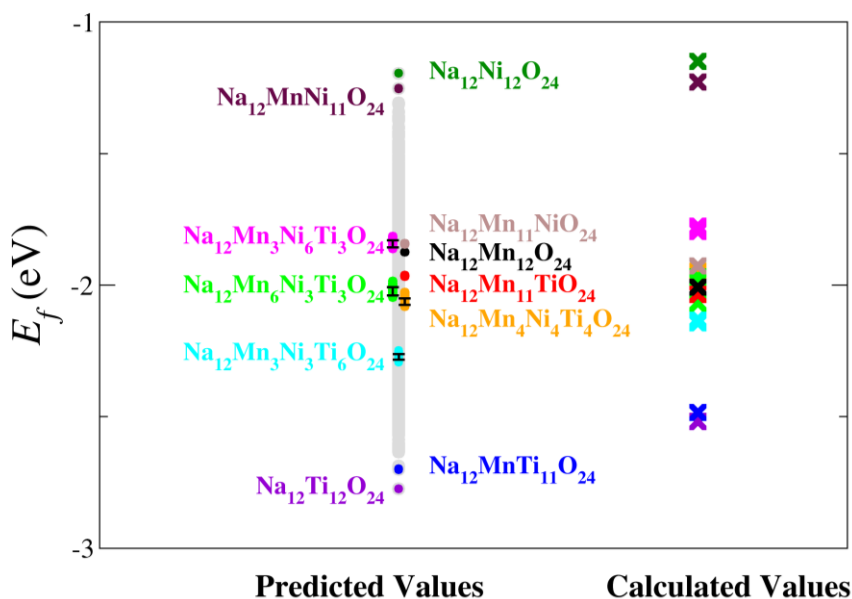


Figura 3. Confronto dell'energia di formazione prevista (cerchi) e calcolata (croci) per sistemi non rilassati. Per i sistemi $\text{Na}_{12}\text{Mn}_3\text{Ni}_6\text{Ti}_3\text{O}_{24}$, $\text{Na}_{12}\text{Mn}_6\text{Ni}_3\text{Ti}_3\text{O}_{24}$, $\text{Na}_{12}\text{Mn}_4\text{Ni}_4\text{Ti}_4\text{O}_{24}$ e $\text{Na}_{12}\text{Mn}_3\text{Ni}_3\text{Ti}_6\text{O}_{24}$, l'incertezza statistica delle energie di formazione previste è indicata da linee continue nere centrate nei valori medi e limitate dalle deviazioni standard.

Per valutare ulteriormente le prestazioni predittive del modello di apprendimento automatico, le energie di formazione previste sono state confrontate con quelle calcolate ottenute attraverso il metodo DFT gestito dal flusso di lavoro AiIDA. In questa seconda fase, la E_f calcolata in approccio DFT è relativa ai sistemi rilassati. La Figura 4 mostra i grafici di dispersione delle energie di formazione previste rispetto a quelle calcolate per due serie di sistemi: sistemi non rilassati (cerchi rossi) sul pannello (a); sistemi rilassati (cerchi blu) sul pannello (b). Ogni punto dati rappresenta una configurazione univoca della composizione del materiale. Come illustrato, la maggior parte dei punti si trova vicino alla linea $y = x$ (linea tratteggiata nera), indicando un forte accordo tra le previsioni e i valori calcolati. Il raggruppamento stretto attorno a questa linea dimostra l'accuratezza del modello in un'ampia gamma di energie di formazione.

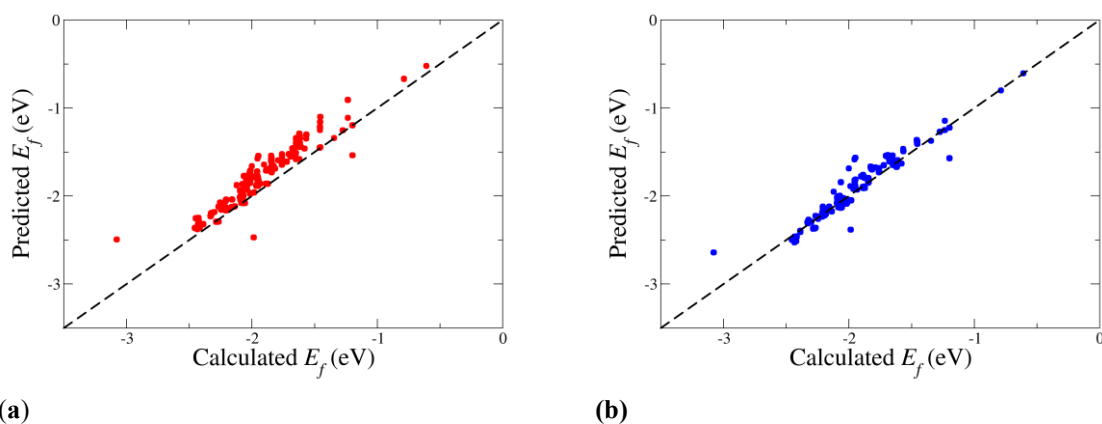


Figura 4. Grafici a dispersione delle energie di formazione previste rispetto a quelle calcolate E_f utilizzando sistemi non rilassati (pannello (a) cerchi rossi) e rilassati (pannello (b) cerchi blu). La linea tratteggiata nera rappresenta la relazione ideale $y = x$.

Tabella 2. L'errore assoluto medio (MAE), il coefficiente di determinazione (R^2) e il coefficiente di correlazione di Pearson (r) sono stati calcolati confrontando le energie di formazione previste e calcolate per i sistemi non rilassati e rilassati.

Sistemi	MAE (eV)	R^2	R di Pearson
Non rilassato	0.166	0.665	0.960
Rilassato	0.052	0.944	0.975

Inoltre, l'accuratezza delle previsioni è stata quantificata utilizzando l'errore assoluto medio (MAE), il coefficiente di determinazione (R^2) e la correlazione r di Pearson, come riportato nella Tabella 2. Dal confronto dei grafici di dispersione, emergono due osservazioni chiave: 1) il GeoCGNN può fornire previsioni ragionevolmente accurate per l'energia di formazione utilizzando sistemi non rilassati direttamente dal file CIF di input; 2) la precisione migliora per i sistemi rilassati, come indicato dall'allineamento più stretto dei punti con la diagonale nel grafico della Figura 4b. La prima osservazione è significativa perché abbiamo addestrato il modello DL a fare previsioni risparmiando tempo, bypassando la necessità di costosi calcoli computazionali a principi primi. La struttura rilassata si ottiene solo dopo che il calcolo è completo (con entrambe l'energia di formazione e la struttura rilassata come output). Per quanto ne sappiamo, attualmente non esistono modelli DL in grado di prevedere la geometria del sistema rilassato. Pertanto, l'approccio standard consiste nel prevedere l'energia di formazione della struttura cristallina non rilassata. I nostri risultati mostrano che il GeoCGNN raggiunge una precisione soddisfacente in questo compito. La seconda osservazione è altrettanto importante, in quanto le migliori prestazioni degli algoritmi predittivi su sistemi rilassati concordano con il fatto che Quantum Espresso calcola l'energia di formazione delle strutture rilassate. Questa è la conferma che l'addestramento della GeoCGNN ha avuto successo.

11.6 Algoritmi predittivi per misure sperimentali

In questa sezione mostriamo come un'analisi con algoritmi ML di un dataset che esplori le variazioni chimico-strutturali e di prestazioni delle batterie in funzione delle quantità relative di metalli di transizione negli ossidi misti a struttura laminare può essere un processo complesso ma molto utile per l'ottimizzazione delle prestazioni delle batterie. Abbiamo reso disponibile nel repository <https://gitlab.brindisi.enea.it/claudio.ronchetti/ai4mat> il codice completo per addestrare e valutare diversi modelli di regressione.

Il dataset può includere i seguenti parametri:

1. *Composizione dei Metalli di Transizione*: Percentuali di ciascun metallo di transizione.
2. *Struttura Cristallina*: Parametri della cella unità, angoli, e volume.
3. *Morfologia*: Dimensioni delle particelle, superficie specifica.
4. *Proprietà Elettriche*: Conducibilità elettrica, resistenza interna.
5. *Prestazioni della Batteria*: Capacità specifica, densità energetica, efficienza coulombica, cicli di carica/scarica.
6. *Proprietà Chimiche*: Stabilità termica, stabilità chimica, tasso di ossidazione/riduzione.

Ad esempio, un dataset potrebbe elencare le seguenti proprietà, rappresentabili in colonne di una tabella:

- Sample ID: Identificatore univoco del campione.

- Mn (%), Ni (%), Co (%), Cu (%): Percentuali in peso di ciascun metallo di transizione.
- a, b, c (Å): Lunghezze dei lati della cella unità.
- α , β , γ (°): Angoli tra i lati della cella unità.
- Particle Size (nm): Dimensioni medie delle particelle.
- Surface Area (m²/g): Superficie specifica del materiale.
- Electrical Conductivity (S/cm): Conducibilità elettrica del materiale.
- Internal Resistance (Ω): Resistenza interna del materiale quando utilizzato in una batteria.
- Specific Capacity (mAh/g): Capacità specifica della batteria.
- Energy Density (Wh/kg): Densità energetica della batteria.
- Coulombic Efficiency (%): Efficienza coulombica della batteria.
- Charge/Discharge Cycles: Numero di cicli di carica/scarica che la batteria può sostenere prima di un significativo calo di prestazioni.
- Thermal Stability (°C): Temperatura alla quale il materiale rimane stabile.

I dati mostrano come le diverse proprietà chimiche e fisiche variano con la composizione, ovvero variando le percentuali dei metalli di transizione. La caratterizzazione sperimentale impiega tecniche come XRD per la struttura cristallina, SEM/TEM per la morfologia, e misure di conducibilità elettrica e resistenza interna. Inoltre, nel dataset vengono valutate le prestazioni elettrochimiche attraverso test di cicli di carica/scarica e misure di capacità e densità energetica. Questo dataset può essere utilizzato per analisi statistiche, modellazione e simulazioni per ottimizzare la composizione e le condizioni di fabbricazione degli ossidi misti per migliorare le prestazioni delle batterie.

Per costruire un modello di machine learning che predica le proprietà delle batterie in base alla composizione dei metalli di transizione, possiamo usare un algoritmo di regressione. Abbiamo utilizzato la libreria Python scikit-learn per questo scopo.

Supponendo che il dataset sia memorizzato in un file CSV, il codice per analisi del dataset è strutturato come segue:

1. *Caricamento dei Dati*: I dati vengono caricati da un file CSV usando *Pandas*.
2. *Selezione delle Caratteristiche e del Target*: Le colonne delle caratteristiche e del target vengono separate.
3. *Divisione del Dataset*: Il dataset viene diviso in set di addestramento e di test.
4. *Standardizzazione delle Caratteristiche*: Le caratteristiche vengono standardizzate per avere media 0 e deviazione standard 1.
5. *Addestramento del Modello*: Viene utilizzato un modello di regressione, ad esempio a foresta casuale (*Random Forest Regressor*) per addestrare i dati.
6. *Valutazione del Modello*: Viene calcolato l'errore quadratico medio (MSE) e il coefficiente di determinazione (R^2) per valutare le prestazioni del modello.
7. *Predizioni*: Viene creata una funzione per fare nuove previsioni basate su nuovi campioni di dati.

Ovviamente, è possibile aggiungere ulteriori pre-processing e ottimizzazione del modello per migliorare le prestazioni predittive.

Esistono diversi modelli di regressione che possono essere utilizzati per predire le proprietà delle batterie in base alla composizione dei metalli di transizione. Ognuno di questi modelli ha i suoi vantaggi e svantaggi, e la scelta del modello dipende dalle caratteristiche specifiche del dataset e dal problema. Ecco alcuni modelli di regressione che abbiamo considerato:

1. *Linear Regression*: È un modello di base che assume una relazione lineare tra le caratteristiche e il target.
2. *Ridge Regression*: È una forma di regressione lineare che include una penalità per la grandezza dei coefficienti per evitare overfitting.
3. *Lasso Regression*: Simile a Ridge Regression, ma utilizza una penalità L1 che può portare a coefficienti esattamente zero, risultando in un modello più semplice.
4. *Elastic Net Regression*: Combina le penalità di Ridge e Lasso Regression.
5. *Decision Tree Regressor*: Utilizza un albero decisionale per modellare le relazioni non lineari tra le caratteristiche e il target.
6. *Support Vector Regressor (SVR)*: Utilizza le macchine a vettori di supporto per eseguire la regressione, utile per dataset con piccole dimensioni e molte caratteristiche.
7. *Gradient Boosting Regressor*: Combina la potenza di molteplici modelli deboli (alberi decisionali) per creare un modello potente.
8. *XGBoost*: Un'implementazione efficiente dell'algoritmo di boosting.

Potrebbe essere utile esplorare nuove caratteristiche derivate dalle esistenti realizzando il così detto *feature engineering*. Il codice disponibile in <https://gitlab.brindisi.enea.it/claudio.ronchetti/ai4mat> permette di sperimentare i diversi modelli e le tecniche di ottimizzazione per trovare la combinazione migliore per il dataset da analizzare.

11.7 Conclusioni

Nella presente attività sono stati raggiunti i seguenti risultati:

- Applicazione del workflow di automatizzazione dei calcoli ab initio ai database dei materiali esistenti. I calcoli DFT sono stati automatizzati e integrati in una catena di flusso di lavoro computazionale sviluppata all'interno del framework AiiDA (D1.23) applicato a strutture cristalline estratte dal database Materials Project. L'energia totale dei materiali è stata utilizzata per valutare l'energia di formazione e il potenziale di intercalazione.

- Database ENEA di dati ab initio di materiali catodici basati su ossidi metallici lamellari puri e drogati integrata per accumulatori.

Il set di dati generato in questo studio è disponibile all'indirizzo <https://iemap.enea.it/>. Il workflow AiiDA ha generato fino ad oggi un totale di oltre 700 elaborazioni completate, che sono state memorizzate nel database. Questi dati rappresentano una risorsa fondamentale per l'analisi e l'interpretazione dei risultati.

- Analisi machine learning dei materiali cristallini per catodi di batterie e identificazione di potenziali candidati per la sintesi dei materiali.

Abbiamo adottato un approccio data-driven in cui il GeoCGNN è stato utilizzato per prevedere l'energia di formazione come proprietà target dei cristalli generati dopo l'addestramento su un set di dati estratto dal database Materials Project [41]. L'input del modello GeoCGNN è un grafo integrato con altre informazioni sul reticolo cristallino. Sono stati illustrati la progettazione, il processo di addestramento e gli strumenti computazionali utilizzati per ottenere previsioni accurate sul set di dati. Abbiamo valutato le prestazioni del modello pre-addestrato su un set di dati completamente diverso che non è stato incluso nella fase di addestramento originale. In particolare, l'assessment si concentra sulle batterie a ioni sodio, scelti per le loro proprietà distintive, per determinare fino a che punto il modello pre-addestrato è in grado di generalizzarsi a dati nuovi e inediti. In questo modo, abbiamo convalidato la robustezza e la trasferibilità del modello pre-addestrato, offrendo approfondimenti sulla sua più ampia applicabilità e sui suoi limiti quando applicato a diversi set di dati. Il modello meglio addestrato è stato eseguito su strutture drogate non rilassate di interesse per prevedere le energie di formazione e quindi calcolare i potenziali redox. È stato applicato un metodo multi-obiettivo per rilevare l'optimum utilizzando entrambi i valori delle proprietà dei materiali. Finora, l'optimum si basa sull'energia di formazione e sul potenziale redox, ma possono essere integrati altri fattori, ad esempio la riciclabilità e la scarsità di elementi. Il lavoro è interessante e i risultati del calcolo mostrano che le GNN riducono il tempo necessario per prevedere le proprietà del materiale di un fattore di migliaia rispetto all'originale. I risultati hanno rivelato che il drogaggio del titanio migliora la stabilità del sistema, mentre il drogaggio del nichel la riduce. Per valutare le prestazioni predittive del modello DL, le energie di formazione previste sono state confrontate con le energie di formazione calcolate ottenute dai calcoli DFT. I nostri risultati mostrano che il GeoCGNN raggiunge un'accuratezza soddisfacente nel prevedere l'energia di formazione di strutture cristalline non rilassate, con un'accuratezza che migliora quando i sistemi rilassati vengono utilizzati come input. Ciò dimostra che il GeoCGNN è stato addestrato con successo per prevedere risultati coerenti con i calcoli DFT, che calcolano l'energia di formazione delle strutture rilassate. Tuttavia, poiché le strutture rilassate sono disponibili solo dopo aver eseguito il calcolo dell'energia totale DFT, l'approccio più efficace senza eseguire calcoli DFT consiste nell'utilizzare sistemi non rilassati come input per il GeoCGNN per generare previsioni.

Infine, abbiamo reso disponibile un codice che esegua l'analisi con algoritmi ML di un dataset che esplori le variazioni chimico-strutturali e di prestazioni delle batterie in funzione di quantità relative di metalli di transizione negli ossidi misti a struttura laminare.

Prospettive future e considerazioni finali.

In conclusione, per valutare la stabilità dei materiali catodici NaMnO_2 di ossido a strati P2 drogati in modo sostitutivo, abbiamo adottato due diversi metodi computazionali: 1) calcoli DFT ad alto rendimento ad alto rendimento; e 2) tecniche ML basata sull'addestramento di un GeoCGNN. In entrambi i casi, abbiamo

esplorato un ampio spazio combinatorio di materiali candidati per batterie catodiche generando numerose composizioni alternative attraverso la sostituzione casuale di atomi di Mn.

Per migliorare ulteriormente il modello ML ed espanderne l'applicabilità, la versatilità del modello può essere migliorata incorporando proprietà come la banda proibita, il modulo di massa, il modulo di taglio e il rapporto di Poisson nel set di dati di addestramento. Queste proprietà sono spesso correlate a metriche di prestazione critiche, come la stabilità meccanica e le caratteristiche elettroniche, migliorando così la comprensione del comportamento del materiale. Inoltre, riutilizzando un modello di ML pre-addestrato, come quello sviluppato per prevedere le proprietà elettroniche, potrebbe essere possibile mettere a punto il modello per altre proprietà correlate, come il potenziale redox, utilizzando il transfer learning per le tecniche di previsione tra proprietà. Questo approccio potrebbe ridurre significativamente i requisiti di risorse computazionali, sfruttando al contempo le conoscenze esistenti.

La correzione di Hubbard ha migliorato la descrizione della correlazione degli orbitali d nei gusci elettronici dei metalli di transizione e la stima del gap energetico. Tuttavia, il funzionale di scambio-correlazione meta-GGA r2SCAN [63] si comporta meglio di GGA nel prevedere le proprietà termodinamiche di vari metalli di transizione 3d, 4d e 5d [64]. Pertanto, l'adozione della funzionalità r2SCAN nei flussi di lavoro è una strategia promettente per la ricerca futura.

Come osservazione finale, vale la pena notare che l'esecuzione di un tipico calcolo DFT per una singola struttura cristallina richiede un tempo che va da una a 24 ore, mentre l'addestramento del modello dal set di dati richiede alcune ore e una singola previsione viene effettuata in frazioni di secondo. Ciò dimostra che il GeoCGNN riduce il tempo necessario per prevedere le proprietà dei materiali di diversi ordini di grandezza rispetto ai tradizionali calcoli DFT, pur mantenendo un'accuratezza soddisfacente. Inoltre, la riduzione del consumo di energia è significativa e non deve essere trascurata.

11.8 Riferimenti bibliografici

- [1] M. R. A. Bhuiyan - Overcome the future environmental challenges through sustainable and renewable energy resources. *Micro Nano Lett.* 17 (2022) 402.
- [2] A. Chakraborty, S. Kunnikuruvan, S. Kumar, B. Markovsky, D. Aurbach, M. Dixit, D. T. Major - Layered Cathode Materials for Lithium-Ion Batteries: Review of Computational Studies on $\text{LiNi}_{1-x-y}\text{Co}_x\text{Mn}_y\text{O}_2$ and $\text{LiNi}_{1-x-y}\text{Co}_x\text{Al}_y\text{O}_2$. *Chem. Mater.* 32 (2020) 915.
- [3] S. W. Kim, D. H. Seo, X. Ma, G. Ceder, K. Kang - Electrode Materials for Rechargeable Sodium-Ion Batteries: Potential Alternatives to Current Lithium-Ion Batteries. *Adv. Energy Mater.* 2 (2012) 710.
- [4] R. J. Clement, P. G. Bruce, C. P. Grey - Review—Manganese-Based P2-Type Transition Metal Oxides as Sodium-Ion Battery Cathode Materials. *J. Electrochem. Soc.* 162 (2015) A2589.
- [5] Q. Liu, Z. Hu, M. Chen, C. Zou, H. Jin, S. Wang, S. L. Chou, S. X. Dou - Recent Progress of Layered Transition Metal Oxide Cathodes for Sodium-Ion Batteries. *Small* 15 (2019) 1805381.
- [6] Z. Jiaolong, W. Wenhui, W. Wei, W. Shuwei, L. Baohua - Comprehensive Review of P2-Type $\text{Na}_{2/3}\text{Ni}_{1/3}\text{Mn}_{2/3}\text{O}_2$, a Potential Cathode for Practical Application of Na-Ion Batteries. *ACS Appl. Mater. Interfaces* 11 (2019) 22051.
- [7] Dae Hoe Lee, Jing Xu, Ying Shirley Meng - An advanced cathode for Na-ion batteries with high rate and excellent structural stability. *Phys. Chem. Chem. Phys.* 15 (2013) 3304.
- [8] P. P. Prosini, M. Carewska, C. Cento, G. Tarquini, F. Maroni, A. Birrozzi, F. Nobili - Tin-Decorated Reduced Graphene Oxide and $\text{NaLi}_{0.2}\text{Ni}_{0.25}\text{Mn}_{0.75}\text{O}_6$ as Electrode Materials for Sodium-Ion Batteries. *Materials* 12 (2019) 1074.
- [9] Y. K. Sun, Y. S. Jeon, H. J. Leeb - Overcoming Jahn-Teller Distortion for Spinel Mn Phase. *Electrochem. Solid-State Lett.* 3 (1999) 7.
- [10] M. M. Thackeray - Manganese oxides for lithium batteries. *Prog. Solid State Chem.* 25 (1997) 1.
- [11] P. Gupta, S. Pushpakanth, M. Ali Haider, S. Basu - Understanding the Design of Cathode Materials for Na-Ion Batteries. *ACS Omega* 7 (2022) 5605.

- [12] Thang Phan Nguyen, Il Tae Kim. - Recent Advances in Sodium-Ion Batteries: Cathode Materials. *Materials* 16 (2023) 6869.
- [13] W. Haoji, G. Xu, Z. Shu, M. Yu, N. Lianshan, G. Jinqiang, L. Huanqing, H. Ningyun, Z. Baichao, Z. Fangjun et al. - High-Entropy Na-Deficient Layered Oxides for Sodium-Ion Batteries. *ACS Nano* 17 (2023) 12530.
- [14] H. Yoshida, N. Yabuuchi, K. Kubota, I. Ikeuchi, A. Garsuch, M. Schulz-Dobrick, S. Komaba - P2-Type $\text{Na}_{2/3}\text{Ni}_{1/3}\text{Mn}_{2/3-x}\text{TixO}_2$ as a New Positive Electrode for Higher Energy Na-Ion Batteries. *Chem. Commun.* 50 (2014) 3677.
- [15] C. Zhijie, L. Lijiang, Z. Chaojin, M. Xiaobo, W. Hailong - Contribution of titanium substitution on improving the electrochemical properties of $\text{P2-Na}_{0.67}\text{Ni}_{0.33}\text{Mn}_{0.67}\text{O}_2$ cathode material for sodium-ion storage *Funct. Mater. Lett.* 13 (2020) 2051010.
- [16] T. Lombardo, M. Duquesnoy, H. El-Bouysidy, F. Årén, A. Gallo-Bueno, P. Bjørn Jørgensen, A. Bhowmik, A. Demortière, E. Ayerbe, F. Alcaide et al. - Artificial Intelligence Applied to Battery Research: Hype or Reality? *Chem. Rev.* 122 (2022) 10899.
- [17] Y. Yuan, A. Alabdulkareem, A. S. Pentland - An interpretable approach for social network formation among heterogeneous agents. *Nature Communications* 9 (2018) 4704.
- [18] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, D. Yin - Graph neural networks for social recommendation. In *The world wide web conference* (2019) 417.
- [19] S. L. Rasmussen, C. Pertoldi, D. W. Macdonald - Machine-learning prediction of hosts of novel coronaviruses requires caution as it may affect wildlife conservation. *Nat Commun* 13 (2022) 5101.
- [20] K. Schutt, P. J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko, K. R. Muller - SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems* 30 (2017) 992.
- [21] K. T. Schutt, F. Arbabzadah, S. Chmiela, K. R. A. Mueller, K. A. Tkatchenko - Quantum-chemical insights from deep tensor neural networks. *Nature Communications* 8 (2017) 1.
- [22] T. Xie, J. C. Grossman - Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters* 120 (2018) 145301.
- [23] C. Chen, W. Ye, Y. Zuo, C. Zheng, S. P. Ong - Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials* 31 (2019) 3564.
- [24] C. W. Park, C. Wolverton - Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Physical Review Materials* 4 (2020) 063801.
- [25] Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, T. F. Miller III - OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *The Journal of Chemical Physics* 153 (2020) 124111.
- [26] K. Choudhary, B. DeCost - Atomistic Line Graph Neural Network for improved materials property predictions. *npj Computational Materials* 7 (2021) 1.
- [27] M. L. Pasini, P. Zhang, S. T. Reeve, J. Y. Choi - Multitask graph neural networks for simultaneous prediction of global and atomic properties in ferromagnetic systems. *Machine Learning: Science and Technology* 3 (2022) 025007.
- [28] J. Cheng, C. Zhang, L. Dong - A geometric-information-enhanced crystal graph network for predicting properties of materials. *Communications Materials* 2 (2021) 1.
- [29] S. S. Young, F. Yuan, M. Zhu - Chemical descriptors are more important than learning algorithms for modelling. *Molecular Informatics* 31 (2012) 707.
- [30] P. G. Polishchuk, V. E. Kuzmin, A. G. Artemenko, E. N. Muratov - Universal approach for structural interpretation of QSAR/QSPR models. *Molecular Informatics* 32 (2013) 843.
- [31] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha - Universal fragment descriptors for predicting properties of inorganic crystals. *Nature Communications* 8 (2017) 1.
- [32] J. Schmidt, M. R. Marques, S. Botti, M. A. Marques - Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials* 5 (2019) 1.

- [33] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, S. Chintala - Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).
- [34] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga et al. - Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).
- [35] S. A. Weil, S. A. Brandt, E. L. Miller, D. D. Long, C. Maltzahn - Ceph: A scalable, high-performance distributed file system. In *Proceedings of the 7th symposium on Operating systems design and implementation* (2006) 307.
- [36] T. Xie – CGCNN. <https://github.com/txie-93/cgcnn> (2022).
- [37] J. Cheng, C. Zhang, L. Dong - Geo-CGNN. <https://github.com/Tinystormjojo/geo-CGNN> (2022).
- [38] Sylabs – Singularity. [sylabs.io. https://sylabs.io/singularity/](https://sylabs.io/singularity/) (2022).
- [39] A. Mariano, G. D'Amato, G. Formisano, G. Guarnieri, G. Santomauro, S. Migliori - Crescoware: A Container-Based Gateway for HPC and AI Applications in the ENEAGRID Infrastructure. In *Science and Information Conference*. Springer, Cham (2022) 196.
- [40] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl - Neural message passing for quantum chemistry. *International conference on machine learning* 70 (2017) 1263.
- [41] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder et al. - Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials* 1 (2013) 011002.
- [42] S. I. Amari - Backpropagation and stochastic gradient descent method. *Neurocomputing* 5 (1993) 185.
- [43] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni et al. - Advanced capabilities for materials modelling with Quantum ESPRESSO. *J. Phys. Condens. Matter* 29 (2017) 465901.
- [44] P. Hohenberg, W. Kohn - Inhomogeneous electron gas. *Phys. Rev.* 136 (1964) B864.
- [45] W. Kohn, L. J. Sham - Self-consistent equations including exchange and correlation effects. *Phys. Rev.* 140 (1965) A1133.
- [46] J. P. Perdew, K. Burke, M. Ernzerhof - Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77 (1996) 3865.
- [47] H. J. Monkhorst, J. D. Pack - Special points for Brillouin-zone integrations. *Phys. Rev. B* 13 (1976) 5188.
- [48] N. Marzari, D. Vanderbilt, A. De Vita, M. C. Payne - Thermal contraction and disordering of the Al(110) surface. *Phys. Rev. Lett.* 82 (1999) 3296.
- [49] B. Himmetoglu, A. Floris, S. Gironcoli, M. Cococcioni - Hubbard-corrected DFT energy functionals: The LDA+U description of correlated systems. *Int. J. Quantum Chem.* 114 (2014) 14.
- [50] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, B. Kozinsky - AiiDA: automated interactive infra-structure and database for computational science. *Comput. Mater. Sci.* (2016) 218.
- [51] S. P. Huber, S. Zoupanos, M. Uhrin, L. Talirz, L. Kahle, R. Hauselmann, D. Gresch, T. Müller, A. V. Yakutovich, C. W. Andersen et al. - AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Scientific data* 7 (2020) 300.
- [52] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J. W. Boiten, L. B. Silva Santos, P. E. Bourne et al. - The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3 (2016) 1.
- [53] J. J. de Pablo, N. E. Jackson, M. A. Webb, L. Q. Chen, J. E. Moore, D. Morgan, R. Jacobs, T. Pollock, D. G. Schlom, E. S. Toberer et al. - New frontiers for the materials genome initiative. *npj Computational Materials* 5 (2019) 1.
- [54] C. Draxl, M. Scheffler - NOMAD: The FAIR concept for big data-driven materials science. *Mrs Bulletin* 43 (2018) 676.
- [55] L. Talirz, S. Kumbhar, E. Passaro, A. V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S. P. Huber, S. Zoupanos et al. - Materials Cloud, a platform for open computational science. *Scientific data* 7 (2020) 1.

- [56] A. Belsky, M. Hellenbrandt, V. L. Karen, P. Luksch - New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallographica Section B: Structural Science* 58 (2020) 364.
- [57] F. Iannone, F. Ambrosino, G. Bracco, M. De Rosa, A. Funel, G. Guarnieri, S. Migliori, F. Palombi, G. Ponti, G. Santomauro et al. - CRESCO ENEA HPC clusters: a working example of a multifabric GPFS Spectrum Scale layout. 2019 International Conference on High Performance Computing & Simulation (HPCS), Dublin, Ireland (2019) 1051.
- [58] A. Mariano, G. D'amato, F. Ambrosino, G. Aprea, F. Buonocore, M. Celino, A. Colavincenzo, M. Fina, A. Funel, S. Giusepponi et al. - Fast Access to Remote Objects 2.0 a renewed gateway to ENEAGRID distributed computing resources. *Futur. Gener. Comput. Syst.* 94 (2019) 920.
- [59] M. Harper et al. - python-ternary: Ternary Plots in Python. Zenodo. 10.5281/zenodo.594435. Zenodo (2015).
- [60] I. Goodfellow, Y. Bengio, A. Courville - *Deep Learning*. MIT Press: Cambridge, MA, USA, (2016).
- [61] A. Géron - *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media (2019).
- [62] C. M. Bishop - *Pattern Recognition and Machine Learning*. Springer (2006).
- [63] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, J. Sun - Accurate and Numerically Efficient r2SCAN Meta-Generalized Gradient Approximation. *J. Phys. Chem. Lett.* 11 (2020) 8208.
- [64] M. Kothakonda, A. D. Kaplan, E. B. Isaacs, C. J. Bartel, J. W. Furness, J. Ning, C. Wolverton, J. P. Perdew, J. Sun - Testing the r2SCAN Density Functional for the Thermodynamic Stability of Solids with and without a van der Waals Correction. *ACS Mater. Au* 3 (2023) 102.

11.9 Abbreviazioni ed acronimi

AI = Artificial Intelligence
 AiiDA = Automated Interactive Infrastructure and Database for computational science
 ALIGNN = Atomistic Line Graph Neural Network
 BFGS = Broyden-Fletcher-Goldfarb-Shanno
 CIF = Crystallographic Information File
 CGCNN = Crystal Graph Convolutional Neural Network
 CNN = Convolutional Neural Network
 CSV = Comma-Separated Values
 DFT = Density Functional Theory
 DFT+U = Teoria del funzionale della densità di carica con il parametro Hubbard U
 FAIR (dati) = Dati che rispettano i principi di rintracciabilità (findability), accessibilità (accessibility), interoperabilità (interoperability) e riusabilità (reusability)
 FC = Fully Connected
 GeoCGNN = Geometric-information-enhanced Crystal Graph Neural Network
 GGA = General Gradient Approximation
 GGA r2SCAN = GGA con SCAN revisionato
 GNN = Graph Neural Network
 HPC = High Performance Computing
 Hydra GNN = rete neurale grafica multi-task
 iCGCNN = CGCNN che incorpora l'informazione della struttura cristallina tassellata di Voronoi
 LA = Linea di attività
 LIB = batteria a ioni di litio
 LTMO = Layered Transition Metal Oxide
 M = metallo
 MAE = Mean Average Error

MEGNet = MatErials Graph Network
ML = Machine Learning
MLP = Multilayer Perceptron
MP = Materials Project
MPNN = Message Passing Neural Network
MSE = Mean Square Error
NIB = Na Ion Battery (batteria a ioni di sodio)
NN = Neural Network
NOMAD = Novel Materials Discovery
PBC = Periodic Boundary Conditions (Condizioni periodiche al contorno)
PBE = Funzionale di scambio correlazione di Perdew–Burke–Ernzerhof
PLMF = Property-Labeled Material Fragments
PP = Pseudo-Potential
PW = Plane-Wave
PWSCF = Plane-Wave Self-Consistent Field
QE = Quantum ESPRESSO (software)
SCAN = Strongly Constrained and Appropriately Normed
SCF = Self-Consistent Field
SchNet = rete neurale convoluzionale a filtro continuo
SEM = scanning electron microscopy
SGD = Stochastic Gradient Descent
TEM = transmission electron microscopy
TM = Transition Metal
XRD = X-Ray Diffraction
XGBoost = eXtreme Gradient Boosting