

Classificazione della malattia di Alzheimer utilizzando dati MRI basati su tecniche di Deep Learning

<https://github.com/julianpajo/Alzheimer-Classification>

Julian Pajo

Università degli Studi Aldo Moro

Bari, Italia

j.pajo@studenti.uniba.it

Abstract — L'Imaging a Risonanza Magnetica (MRI) è uno strumento fondamentale per la diagnosi precoce dell'Alzheimer, permettendo di rilevare cambiamenti strutturali e funzionali nel cervello prima dell'insorgenza dei sintomi. Questo articolo propone uno studio approfondito sull'applicazione di tecniche di Deep Learning per l'analisi e la classificazione dei dati MRI. In particolare, viene esplorato il modello di rete neurale convoluzionale (CNN). Il modello proposto comprende le seguenti fasi: pre-processing, che prevede una fase di data augmentation, addestramento del modello di deep learning, e valutazione.

Keywords— *Alzheimer, Imaging a Risonanza Magnetica (MRI), Deep learning (DL), Rete Neurale Convoluzionale (CNN)*

I. INTRODUZIONE

L'Alzheimer rappresenta il 70% dei casi di demenza a livello mondiale, rendendolo il tipo di demenza più diffuso. Si tratta di una condizione neurologica irreversibile che compromette gradualmente le capacità cognitive [1].

La ricerca per la diagnosi automatica della malattia di Alzheimer (AD) basata su tecniche tradizionali di machine learning è in continua evoluzione, e gli approcci basati sul deep learning stanno diventando una scelta sempre più popolare per la diagnosi dell'AD. Le tecniche all'avanguardia che considerano una diagnosi multimodale hanno dimostrato di avere un'accuratezza superiore rispetto alla diagnosi manuale [2].

Pertanto, questa ricerca vuole descrivere una tecnica basata su deep learning per la rilevazione precoce dell'Alzheimer. Lo studio intende verificare l'utilità di questo approccio nella diagnosi dell'Alzheimer in fase iniziale utilizzando dati di risonanza magnetica cerebrale.

II. DATA COLLECTION E PREPROCESSING

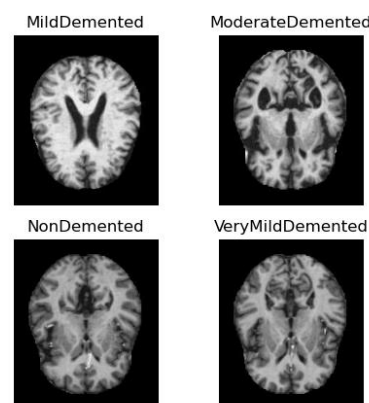
In questa sezione, descriviamo il dataset utilizzato nel nostro studio e discutiamo i passaggi di pre-processing eseguiti sui dati.

II.1 DATASET

La risonanza magnetica (MRI) è la modalità de facto negli studi sul cervello grazie al suo superiore contrasto delle immagini nei tessuti molli senza l'uso di radiazioni ionizzanti. Le immagini MRI sono ampiamente utilizzate anche per esaminare altre regioni anatomiche [3].

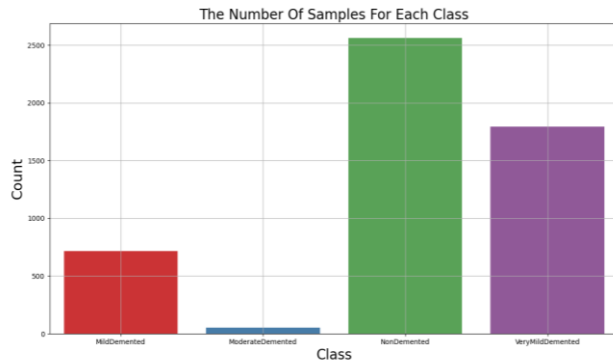
I dati utilizzati nella preparazione di questo articolo sono stati ottenuti da un dataset disponibile su Kaggle denominato "Alzheimer's Dataset 4 Class of Images". Questo dataset contiene immagini di risonanza magnetica (MRI) del cervello di pazienti con quattro diverse classi di Alzheimer: normale, moderata, moderatamente grave e grave. I dati comprendono quattro classi di immagini sia nel set di addestramento che in quello di test, con un totale complessivo di circa 5000 immagini.

Di seguito sono presentate le immagini estratte dal dataset, ognuna associata al relativo livello di Alzheimer.



II.11 PREPROCESSING

Per il preprocessing dei dati, è stata condotta un'analisi preliminare. Durante questa fase è emerso che il numero di immagini per categoria nel set di addestramento non presentava un equilibrio significativo.



L'approccio utilizzato per affrontare questo problema integra la data augmentation con il bilanciamento delle classi tramite SMOTE (Synthetic Minority Over-sampling Technique).

La data augmentation è implementata utilizzando ImageDataGenerator, che genera nuove immagini trasformate a partire dal dataset originale. Questo processo espande la dimensione e la varietà del dataset di addestramento attraverso diverse trasformazioni casuali, tra cui normalizzazione, spostamenti orizzontali e verticali, taglio, zoom, rotazioni e inversioni orizzontali. Successivamente, per affrontare il problema dell'imbalance delle classi, i dati aumentati vengono passati all'algoritmo SMOTE. Questo algoritmo genera esempi sintetici delle classi minoritarie per bilanciare il dataset.

La pipeline è come segue:

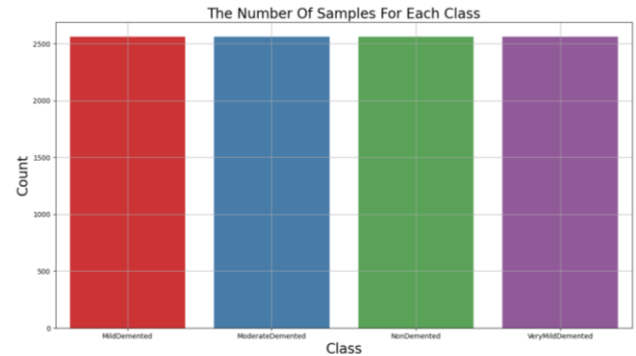
- I dati generati da ImageDataGenerator vengono raccolti e rimodellati in una forma bidimensionale (vettori piatti).
- SMOTE viene applicato a questi dati, generando nuovi esempi sintetici per le classi meno rappresentate.
- I dati sintetizzati vengono infine rimodellati nella loro forma tridimensionale originale, rendendoli pronti per l'addestramento del modello.

Sulle immagini viene applicato un reshape a una dimensione di 128x128 per il train del modello.

Questo processo combinato aumenta la varietà dei dati di addestramento e bilancia la distribuzione delle classi, migliorando la capacità del modello di generalizzare e fornire previsioni più accurate.

A seguito del bilanciamento, si è proceduto con la verifica della nuova distribuzione dei dati. L'output generato evidenzia la riuscita del processo di bilanciamento, come dimostrato dalla forma aggiornata dei dati e delle etichette.

Di seguito sono presentati i risultati ottenuti:



III. REVISIONE DELLA LETTERATURA E MODELLI ESISTENTI

In [5], i ricercatori presentano un framework combinato di MLP (multilayer perceptron) e identificazione RNN (rete neurale ricorrente) di questa malattia utilizzando la risonanza magnetica. Inizialmente, viene utilizzato MLP e successivamente gli autori utilizzano il RNN a 2 livelli formato sul MLP. Hanno raggiunto un'accuratezza del 89,7% per la classificazione dell'AD. Per tutti questi esperimenti, utilizzano i dataset ADNI. Wang Yan et al. [6] utilizzano soggetti a tre classi con dimensioni campione bilanciate e hanno ottenuto una maggiore precisione combinando la risonanza magnetica multimodale con il nucleo CNN. Hanno raggiunto un'accuratezza del 92,06% per questo esperimento. Gunawardena et al. [7] hanno utilizzato un totale di 1615 scansioni e hanno raggiunto un'accuratezza dell'84,4% utilizzando il vettore portante e poi hanno proposto l'approccio CNN ottenendo un'accuratezza del 96%. Ogni tecnica considerata per la rilevazione dell'AD è discussa e i relativi articoli rappresentativi sono brevemente descritti.

IV. MODELLO DI DEEP LEARNING

IV.1 ARCHITETTURA DEL MODELLO

Il modello di deep learning utilizzato in questo studio è una rete neurale convoluzionale (CNN).

Una rete neurale convoluzionale è un tipo di rete neurale artificiale progettata per elaborare dati che hanno una griglia come struttura, come le immagini. È una classe speciale di reti neurali deep learning che sono particolarmente efficaci per l'analisi visiva e possono identificare pattern nelle immagini per riconoscere oggetti, classi e categorie [8].

Una CNN è composta da diversi strati che includono:

- *Strati convoluzionali*: che applicano un insieme di filtri ai dati di input per creare mappe di caratteristiche.
- *Strati di pooling*: che riducono la dimensione spaziale delle mappe di caratteristiche, mantenendo le informazioni più rilevanti.
- *Strati completamente connessi*: che prendono le caratteristiche estratte dagli strati precedenti e le usano per classificare l'input in categorie.

Il modello di deep learning utilizzato in questo studio è una rete neurale convoluzionale (CNN) pre-trainata. Il modello in questione è il VGG19. Il VGG19 è un modello sviluppato dal Visual Geometry Group (VGG) dell'Università di Oxford. È noto per la sua profondità e per essere uno dei modelli di riferimento nell'ambito del riconoscimento e della classificazione delle immagini.

Il VGG19 è composto da 19 livelli, di cui 16 sono strati convoluzionali e 3 sono strati completamente connessi. L'architettura è strutturata come segue:

- *Strati Convoluzionali*: Ci sono in totale 16 strati convoluzionali nel VGG19. Questi strati utilizzano filtri di dimensione (3 \times 3) con uno stride di 1 pixel e padding 'same' per preservare le dimensioni spaziali dell'immagine attraverso gli strati.
- *Strati di Pooling*: Dopo alcuni strati convoluzionali, ci sono strati di pooling massimo ((2 \times 2)) che riducono le dimensioni spaziali dell'output di un fattore di 2.
- *Strati Completamente Connessi*: Alla fine della rete, ci sono 3 strati completamente connessi. Il primo e il secondo hanno ciascuno 4096 canali, mentre l'ultimo strato, che funge da strato di output, ha 1000 canali (per le 1000 classi di ImageNet).
- *Funzioni di Attivazione*: Tutti gli strati nascosti utilizzano la funzione di attivazione ReLU. [9]

Per adattare il modello alle specifiche esigenze del caso di studio e per mitigare il fenomeno dell'overfitting, è stato utilizzato un approccio che combina la rete di base VGG con strati personalizzati. In particolare, inizialmente è stato effettuato il congelamento dei pesi: i primi 15 strati della rete VGG originale sono stati impostati come non allenabili (`trainable=False`), preservando le caratteristiche estratte dalla VGG pre-addestrata e riducendo il rischio di overfitting ai nuovi dati. Gli ultimi 5 strati della VGG sono stati impostati invece come allenabili (`trainable=True`), consentendo l'adattamento fine della rete alle caratteristiche specifiche del nuovo dataset. Successivamente al modello sono stati aggiunti strati personalizzati per migliorarne la capacità di rappresentazione.

Gli strati aggiuntivi sono i seguenti:

- Uno strato di Flatten trasforma le feature map in un vettore piatto.
- Un denso con 512 neuroni, attivazione ReLU, e regolarizzazione L2, per catturare relazioni non lineari tra le caratteristiche.
- Un Batch Normalization per stabilizzare e accelerare l'allenamento, migliorando la robustezza del modello.
- Un Dropout del 50%, per prevenire l'overfitting disattivando casualmente alcuni neuroni durante l'addestramento.
- Un denso con 256 neuroni, attivazione ReLU, e regolarizzazione L2, per una maggiore capacità di rappresentazione.
- Un secondo Batch Normalization e Dropout come il precedente.

Infine, lo strato di output è stato configurato con una funzione di attivazione 'softmax' e 4 neuroni, corrispondenti alle quattro classi del problema di classificazione in esame. La funzione 'softmax' è stata scelta per la sua capacità di fornire una distribuzione di probabilità sulle classi di output, facilitando l'interpretazione dei risultati del modello.

IV.2 ADDESTRAMENTO DEL MODELLO

L'addestramento del modello di deep learning è stato eseguito su una Graphical Processing Unit (GPU). È stata effettuata questa scelta in quanto le GPU sono progettate per gestire molteplici operazioni in parallelo, il che le rende particolarmente adatte per l'addestramento delle CNN che richiedono l'elaborazione simultanea di grandi quantità di dati, come i pixel delle immagini [10]. Si è fatto uso di Conda come sistema di gestione dei pacchetti e dell'ambiente. Il framework principale utilizzato per lo sviluppo e l'addestramento del modello è stato TensorFlow

La GPU utilizzata è stata la NVIDIA RTX 3060. La RTX 3060 è stata supportata da CUDA, una piattaforma di calcolo parallelo e un modello di programmazione sviluppato da NVIDIA. CUDA ha permesso di aumentare significativamente la velocità di addestramento sfruttando i core della GPU.

Inoltre, per ottimizzare le operazioni di rete neurale, è stato impiegato cuDNN, una libreria di primitive per il deep learning altamente ottimizzata per le architetture NVIDIA.

Nella fase di addestramento del modello, il dataset è stato suddiviso in tre sottoinsiemi: il set di addestramento, il set di test e il set di validazione seguendo queste percentuali:

- *Set di Addestramento*: Il 70% dei dati è stato mantenuto come set di addestramento.
- *Set di Test e Validazione*: Il rimanente 30% è stato ulteriormente diviso a metà, con il 15% dei dati originali usato come set di test e il 15% come set di validazione.

Per l'addestramento del modello, è stata impiegata la tecnica di early stopping per prevenire l'overfitting. L'early stopping interrompe l'addestramento quando la metrica monitorata, in questo caso la validation loss, non migliora per un numero consecutivo di epoche definito dalla patience.

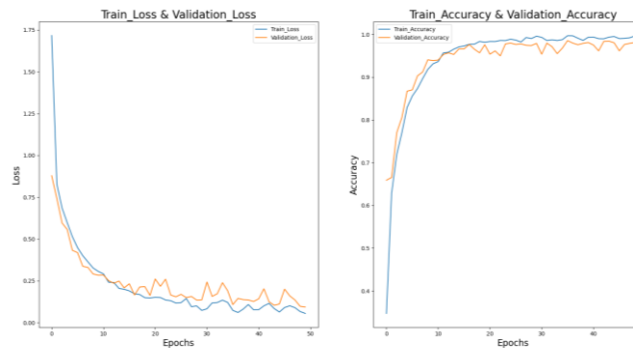
Per addestrare il modello in modo efficiente e stabile, è stato implementato un schema di decadimento esponenziale del tasso di apprendimento. Questo approccio consente di iniziare l'addestramento con un tasso di apprendimento relativamente alto, che facilita grandi aggiornamenti iniziali dei pesi del modello. Successivamente, il tasso di apprendimento diminuisce gradualmente, permettendo al modello di effettuare aggiustamenti più fini man mano che l'addestramento progredisce. Questo schema di decadimento esponenziale verrà poi assegnato al learning rate dell'ottimizzatore 'adam', con la quale il modello verrà compilato insieme alla funzione di perdita 'categorical_crossentropy', che è comune per i problemi di classificazione multi-classe. La metrica di 'accuracy' è stata utilizzata per monitorare le prestazioni del modello.

L'addestramento è stato eseguito per un massimo di 50 epoche e un batch size di 32, con i dati di validazione forniti per valutare le prestazioni del modello ad ogni epoca.

V. SPERIMENTAZIONE E RISULTATI

Durante la fase di valutazione (evaluate), il modello ha mostrato un Test Loss di **0.5748** e un Test Accuracy dell'**82.16%**. Questi risultati sono stati ottenuti utilizzando il set di test che è stato separato dal set di addestramento. Un'accuracy superiore all'80% indica che il modello ha una buona capacità di generalizzazione e predice correttamente la maggior parte delle etichette.

Di seguito viene mostrato l'andamento della loss e dell'accuracy con l'avanzamento delle epoche:



Per quanto riguarda il predict su un set di test completamente nuovo e mai visto prima, l'accuracy scende al **47.22%**. Questa discrepanza tra i due risultati potrebbe essere dovuta a diversi fattori, come la possibile presenza di overfitting o la differenza nella distribuzione dei dati tra i due set di test.

Infatti, come si è potuto osservare dal report di classificazione riportato di seguito, che ci fornisce una visione dettagliata delle prestazioni del modello per ogni classe, includendo la precisione (la percentuale di identificazioni positive che sono effettivamente corrette), il richiamo (la percentuale di casi positivi reali che sono stati identificati correttamente), e l'F1-score (una media armonica tra precisione e richiamo), si nota una notevole discrepanza nel numero di immagini per ciascuna classe nel set di test.

Nel caso in questione, la classe 'NonDemented' ha il maggior numero di immagini (640) e mostra anche la migliore performance in termini di precisione (0.50) e richiamo (0.62). Al contrario, la classe 'ModerateDemented' ha solo 12 immagini e ha ottenuto una precisione e un richiamo di 0.00, indicando che il modello non è riuscito a identificare correttamente alcuna istanza di questa classe.

	precision	recall	f1-score	support
MildDemented	0.14	0.01	0.01	179
ModerateDemented	0.00	0.00	0.00	12
NonDemented	0.50	0.61	0.62	640
VeryMildDemented	0.35	0.21	0.21	448
accuracy			0.40	1279
macro avg	0.25	0.25	0.21	1279
weighted avg	0.39	0.40	0.39	1279

Classification Report

VI. DISCUSSIONE

Una delle principali limitazioni incontrate durante questo studio è stata la dimensione e la completezza del dataset. Un dataset non particolarmente vasto o variegato può limitare la capacità del modello di apprendere caratteristiche generalizzabili e influenzare negativamente la sua performance su dati non visti. Inoltre, una distribuzione non uniforme delle classi può portare a un modello sbilanciato, che tende a favorire le classi maggioritarie a discapito di quelle minoritarie.

Un'altra sfida significativa è stata rappresentata dalle limitazioni della memoria della GPU. Nonostante l'uso di una GPU NVIDIA RTX 3060 abbia notevolmente accelerato il processo di addestramento, la memoria disponibile ha imposto restrizioni sulla dimensione delle immagini in input e sulla complessità dell'architettura del modello. Infatti è la scelta del ridimensionamento delle immagini a 128x128, al posto di usare immagini di più grandi dimensioni come 224x224 ad esempio, è stata dettata dalla limitata disponibilità di memoria.

Questa riduzione comporta pertanto una perdita di dettagli, influenzando la capacità del modello di apprendere caratteristiche complesse e di generalizzare su dati non visti.

Queste modifiche possono avere impatti diretti sulla capacità del modello di catturare dettagli fini e apprendere rappresentazioni complesse dei dati, potenzialmente limitando le sue prestazioni.

Queste limitazioni sottolineano l'importanza di disporre di risorse hardware adeguate e di dataset ampi e ben bilanciati per l'addestramento di modelli di deep learning. Riflessioni future potrebbero includere l'esplorazione di metodi per ottimizzare l'uso della memoria della GPU o l'acquisizione di dataset più ampi per migliorare ulteriormente la robustezza e l'accuratezza del modello.

VII. CONCLUSIONI E PROSPETTIVE FUTURE

Nell'imaging medico, il progresso dell'intelligenza computazionale, dell'apprendimento profondo e del rilevamento assistito dal computer ha attratto molta attenzione nella segmentazione del cervello e nella diagnosi della malattia di Alzheimer (AD). Sebbene i metodi di deep learning abbiano un forte impatto sull'analisi quantitativa delle risonanze magnetiche (MRI) del cervello, è ancora difficile trovare un metodo robusto e generico. Anche se le tecniche di deep learning nell'analisi delle risonanze magnetiche (MRI) cerebrali hanno un grande potenziale, esistono ancora alcune limitazioni. Non mostrano risultati competitivi su dataset relativamente piccoli. In futuro, tuttavia, l'ipotetico incremento della potenza computazionale e l'accessibilità a dataset più ampi e diversificati potrebbero contribuire a superare le attuali limitazioni, rendendo le tecniche di deep learning ancora più preziose nell'analisi delle MRI cerebrali e nella diagnosi di condizioni come l'AD. Se confermato in contesti clinici, questo approccio ha il potenziale per ampliare il campo di applicazione delle tecniche di neuroimaging per la rilevazione e la gestione delle malattie. Ulteriori validazioni potrebbero portare a una cura migliorata e a risultati più favorevoli rispetto all'attuale valutazione neurologica, mentre continua la ricerca di terapie che modificano la malattia.

Questo potrebbe avere un impatto significativo sul modo in cui le condizioni neurologiche sono monitorate e trattate, offrendo ai professionisti della salute strumenti più precisi e una comprensione più profonda delle malattie per personalizzare i trattamenti e migliorare la qualità della vita dei pazienti.

RIFERIMENTI

- [1] Shaymaa E. Sorour, Amr A. Abd El-Mageed, Khalied M. Albarrak, Abdulrahman K. Alnaim, Abeer A. Wafa, Engy El-Shafey Classification of Alzheimer's disease using MRI data based on Deep Learning Techniques (January 2024)
- [2] Samsuddin Ahmed; Kyu Yeong Choi; Jang Jae Lee; Byeong C. Kim; Goo-Rak Kwon; Kun Ho Lee; Ho Yub Jung Ensembles of Patch-Based Classifiers for Diagnosis of Alzheimer Diseases (May 2019)
- [3] M. Cercignani, N. G. Dowell and P. S. Tofts, Quantitative MRI of the Brain: Principles of Physical Measurement, New York, NY, USA: CRC Press, 2018.
- [4] <https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/>
- [5] R. Cui, M. Liu, and G. Li, "Longitudinal analysis for Alzheimer's disease diagnosis using RNN," in Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, 2018, pp. 1398-1401. <https://doi.org/10.1016/j.media.2020.101694>
- [6] Y. Wang, Y. Yang, X. Guo, C. Ye, N. Gao, Y. Fang, et al., "A Novel Multimodal MRI Analysis for Alzheimer's Disease Based on Convolutional Neural Network," in 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 754-757. DOI: 10.1109/EMBC.2018.8512372
- [7] A. Giersch and J. T. Coull, "TRF1: It Was the Best of Time (s)...," Timing & Time Perception, vol. 6, pp. 231-414, 2018. <https://doi.org/10.1163/22134468-00603001>
- [8] <https://www.ibm.com/it-it/topics/convolutional-neural-networks>
- [9] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556.
- [10] <https://ichi.pro/it/convolutional-neural-network-cnn-e-la-sua-applicazione-tutto-quello-che-devi-sapere-266752320713701>