**Rendere DenseNet interpretabile: un caso di studio in radiologia clinica**

**Autori**

* Kwun Ho Ngan (City, University of London, UK)
* Artur d’Avila Garcez (City, University of London, UK)
* Karen M. Knapp (University of Exeter, UK)
* Andy Appelboam (Royal Devon and Exeter Hospital, UK)
* Constantino Carlos Reyes-Aldasoro (City, University of London, UK)

**Riassunto**

L’analisi delle immagini mediche è un compito ripetitivo che, sotto pressione temporale, può portare all’affaticamento e a errori interpretativi. Sebbene i modelli avanzati di deep learning abbiano superato le capacità umane in alcune attività di visione artificiale, la loro adozione in ambito medico è stata limitata da problemi di interpretabilità e dalla mancanza di dati di alta qualità. Questo studio propone un framework per l'interpretazione e la visualizzazione del processo di estrazione delle caratteristiche in una rete neurale convoluzionale profonda. Il framework è applicato alla rilevazione di anomalie nel dataset di radiografie muscolo-scheletriche MURA (Stanford). L'obiettivo è fornire maggiore trasparenza nel funzionamento delle reti DenseNet, migliorando la comprensione del percorso di generazione delle caratteristiche e della logica decisionale.

**Introduzione**

Le radiografie sono uno degli strumenti diagnostici più utilizzati in ambito medico, impiegati per rilevare fratture, anomalie articolari, degenerazioni ossee e altre patologie. Tuttavia, la loro interpretazione è soggetta a errori umani a causa della fatica e del carico di lavoro elevato a cui sono sottoposti i radiologi.

I modelli di deep learning hanno mostrato un enorme potenziale nell’analisi automatizzata delle immagini mediche. Reti neurali convoluzionali avanzate, come DenseNet, sono state applicate con successo a numerosi compiti di visione artificiale, tra cui la classificazione di immagini e il riconoscimento di anomalie. Tuttavia, la scarsa interpretabilità di tali modelli rappresenta un ostacolo significativo alla loro integrazione in ambito clinico.

Per affrontare questa problematica, il presente studio propone un framework di interpretazione dei modelli DenseNet applicati alla radiologia. L'obiettivo è migliorare la trasparenza delle decisioni del modello attraverso tecniche di visualizzazione delle attivazioni, permettendo così ai medici di comprendere meglio il processo di classificazione.

**Background e lavori correlati**

DenseNet è un'architettura di rete neurale convoluzionale che connette ogni livello con tutti i livelli successivi, migliorando la propagazione delle informazioni e riducendo il numero di parametri rispetto a modelli più tradizionali come ResNet. Grazie a questa struttura, DenseNet è in grado di riutilizzare le caratteristiche apprese nei livelli precedenti, rendendo il modello più efficiente e accurato.

Il dataset MURA (Musculoskeletal Radiographs), sviluppato dalla Stanford University, è uno dei più grandi dataset pubblici disponibili per la classificazione di anomalie radiografiche. Contiene oltre 40.000 immagini radiografiche provenienti da più di 12.000 pazienti, classificate come normali o anormali da radiologi esperti.

Studi precedenti hanno dimostrato che i modelli basati su DenseNet possono raggiungere performance comparabili o superiori a quelle umane in alcuni compiti di classificazione. Tuttavia, la loro efficacia varia a seconda del distretto anatomico analizzato e della qualità dei dati di addestramento disponibili.

**Framework proposto**

Il framework proposto utilizza un modello DenseNet-169 addestrato sul dataset MURA per la rilevazione di anomalie radiografiche. Per migliorare la comprensione del modello, vengono applicate tecniche di interpretabilità basate sulla visualizzazione delle attivazioni di livello e sulle mappe di attivazione per classe (Class Activation Maps, CAM).

**Struttura del framework:**

1. **Pre-elaborazione delle immagini:** le immagini vengono ridimensionate a 320x320 pixel e normalizzate per garantire coerenza nei dati di input.
2. **Addestramento del modello:** il modello viene inizializzato con pesi pre-addestrati su ImageNet e successivamente affinato sul dataset MURA. Sono state sperimentate diverse configurazioni di addestramento:
   * Modello pre-addestrato su tutte le categorie anatomiche presenti in MURA.
   * Modello specializzato su specifiche regioni anatomiche (es. polso, dita).
   * Modello addestrato da zero senza pesi pre-addestrati.
3. **Interpretazione delle caratteristiche:** vengono generate visualizzazioni delle attivazioni nei diversi livelli della rete, evidenziando le regioni delle immagini che influenzano maggiormente la classificazione del modello.

**Risultati sperimentali**

Il framework è stato applicato a diverse tipologie di immagini radiografiche per valutare l’interpretabilità del modello. Alcuni risultati rilevanti includono:

* **Frattura del polso:** il modello ha seguito un percorso logico di estrazione delle caratteristiche, identificando inizialmente la struttura del polso e successivamente affinando la localizzazione della frattura.
* **Protesi metalliche:** il modello ha correttamente identificato la presenza di impianti metallici, ma in alcuni casi ha attribuito erroneamente importanza alle etichette testuali presenti nelle immagini.
* **Malattie degenerative articolari:** il modello ha mostrato difficoltà nel rilevare anomalie sottili come degenerazioni articolari lievi, suggerendo che l’addestramento con immagini più dettagliate potrebbe migliorare la performance.
* **Effetto del pre-addestramento:** l’uso di pesi pre-addestrati su ImageNet ha migliorato la capacità del modello di distinguere le strutture anatomiche, riducendo gli errori di classificazione dovuti a informazioni irrilevanti.

**Discussione**

L’analisi delle attivazioni ha fornito preziose informazioni sulle capacità e sui limiti del modello DenseNet. I principali risultati ottenuti includono:

* **Il modello segue un processo di estrazione gerarchico, simile a quello umano.**
* **L’uso di pesi pre-addestrati migliora significativamente la capacità di identificazione delle strutture anatomiche.**
* **Elementi di disturbo come etichette testuali e cornici nelle immagini possono influenzare negativamente la classificazione.**
* **Alcune patologie richiedono una maggiore quantità di dati di addestramento per migliorare il riconoscimento.**

Questi risultati evidenziano l’importanza dell’interpretabilità nella valutazione dei modelli di deep learning per l’ambito medico.

**Conclusione**

Il framework proposto offre un metodo efficace per analizzare e migliorare la trasparenza dei modelli DenseNet applicati alla radiologia. La visualizzazione delle attivazioni aiuta a comprendere il processo decisionale del modello, identificando eventuali bias e problemi di generalizzazione.

L’integrazione di tecniche avanzate di interpretabilità, come i meccanismi di attenzione, potrebbe migliorare ulteriormente la capacità del modello di localizzare anomalie rilevanti, aumentando la fiducia dei clinici nell’uso dell’intelligenza artificiale nella pratica medica.

**Classificazione e Visualizzazione delle Radiografie Normali e Anormali: Un Confronto tra Undici Architetture di Reti Neurali Convoluzionali**

**Autori**

* Ananda Ananda, Kwun Ho Ngan, Cefa Karabağ, Aram Ter-Sarkisov, Eduardo Alonso, Constantino Carlos Reyes-Aldasoro

**Riassunto**

Questo studio esplora la classificazione delle immagini radiografiche utilizzando undici diverse architetture di reti neurali convoluzionali (CNN). Le CNN sono state impiegate per analizzare immagini radiografiche del polso dal dataset MURA di Stanford, suddividendole in due classi: normali e anormali. Sono stati confrontati diversi parametri di ottimizzazione in termini di accuratezza e coefficiente di Cohen Kappa. I due modelli migliori sono stati successivamente sottoposti a tecniche di data augmentation, migliorando significativamente i risultati. Infine, sono state applicate mappe di attivazione per analizzare le regioni d’interesse individuate dai modelli nelle immagini radiografiche.

**Introduzione**

Le fratture del polso e dell’avambraccio sono tra le lesioni più comuni, specialmente negli anziani che possono cadere estendendo il braccio per proteggersi. In alcuni casi, i pazienti potrebbero non rendersi conto della gravità della lesione, lasciandola non trattata, il che può comportare complicazioni significative come rigidità articolare, rottura dei tendini o dolore cronico.

Il trattamento tradizionale delle fratture del polso comprende tecniche di immobilizzazione con gesso o interventi chirurgici come la fissazione interna con placche e viti. Tuttavia, la scelta del trattamento dipende dalla gravità della frattura e dalle condizioni del paziente. Le immagini radiografiche sono il principale strumento diagnostico per determinare il tipo di intervento necessario. A causa dell’enorme volume di immagini da analizzare, l’automazione attraverso l’intelligenza artificiale è diventata una soluzione promettente.

Le reti neurali convoluzionali (CNN) hanno rivoluzionato il campo dell’elaborazione delle immagini, dimostrando prestazioni superiori in vari ambiti della diagnostica medica. Tuttavia, la scelta della migliore architettura CNN per la classificazione delle radiografie non è immediata e richiede un’analisi approfondita. Questo studio confronta undici architetture CNN applicate alla classificazione delle immagini di polsi normali e anormali del dataset MURA e utilizza tecniche di interpretabilità per rendere più trasparente il processo decisionale della rete.

**Metodi e Materiali**

**Dataset MURA**

Il dataset MURA (Musculoskeletal Radiographs) è stato creato dalla Stanford University e contiene oltre 40.000 immagini di radiografie muscolo-scheletriche suddivise in sette categorie anatomiche: gomito, dita, avambraccio, mano, omero, spalla e polso. Ogni immagine è stata classificata da radiologi esperti come normale o anormale. Per questo studio, sono state selezionate solo le immagini della regione del polso, in quanto rappresentano una delle categorie più critiche in termini di diagnosi automatizzata.

**Architetture CNN Analizzate**

Le seguenti architetture di CNN sono state confrontate:

* **GoogleNet**: Un modello profondo che utilizza moduli di inception per migliorare l’efficienza computazionale.
* **VGG-19**: Una rete con 19 strati, caratterizzata dall’uso di filtri convoluzionali 3x3 ripetuti.
* **AlexNet**: Una delle prime reti profonde ad ottenere risultati significativi nel riconoscimento delle immagini.
* **SqueezeNet**: Un’architettura compatta con meno parametri rispetto ad AlexNet, ma con prestazioni comparabili.
* **ResNet-18, ResNet-50, ResNet-101**: Reti con connessioni residue per facilitare l’addestramento di reti molto profonde.
* **Inception-v3**: Un miglioramento del modello GoogleNet con ottimizzazione della gestione dei filtri convoluzionali.
* **DenseNet-201**: Rete che collega ogni livello a tutti i livelli precedenti, migliorando il flusso dell’informazione.
* **Inception-ResNet-v2**: Combina le tecniche delle reti Inception e ResNet per un bilanciamento ottimale tra profondità e efficienza.

**Metodologia di Addestramento**

I modelli sono stati addestrati utilizzando il framework MATLAB R2018b con diverse configurazioni di iperparametri. Le immagini sono state ridimensionate per adattarsi ai requisiti di input di ciascuna rete. Il dataset è stato suddiviso in 90% per il training e 10% per il test. Inoltre, è stata implementata una strategia di data augmentation, includendo rotazioni, riflessioni e regolazioni di contrasto per migliorare la capacità di generalizzazione dei modelli.

**Risultati e Discussione**

**Accuratezza e Kappa di Cohen**

Senza data augmentation, il modello con le migliori prestazioni è stato **Inception-ResNet-v2** con un’accuratezza media di **72.3%** e un coefficiente Kappa di **0.506**. Dopo l’applicazione della data augmentation, le prestazioni sono migliorate notevolmente: **Inception-ResNet-v2 ha raggiunto un’accuratezza dell’85.7% e un Kappa di 0.703**.

**Mappe di Attivazione**

Le mappe di attivazione sono state utilizzate per evidenziare le regioni delle radiografie che i modelli considerano più rilevanti per la classificazione. I risultati mostrano che:

* **Inception-ResNet-v2** ha prodotto mappe di attivazione più focalizzate rispetto a **ResNet-50**, suggerendo un’estrazione delle caratteristiche più precisa.
* Alcuni modelli hanno evidenziato regioni irrilevanti delle immagini, suggerendo la necessità di una migliore pre-elaborazione dei dati.

**Limitazioni e Considerazioni Future**

I modelli testati hanno ottenuto risultati inferiori rispetto ai migliori modelli ensemble nella leaderboard MURA. Ulteriori miglioramenti potrebbero includere:

1. **Pre-elaborazione avanzata**: Rimozione di etichette testuali e miglioramenti nel bilanciamento del dataset.
2. **Tecniche di ensemble learning**: L’uso combinato di più modelli per migliorare la precisione della classificazione.
3. **Integrazione con conoscenze anatomiche**: Incorporare informazioni sulla struttura ossea per migliorare la segmentazione e il riconoscimento delle anomalie.

**Conclusioni**

Questo studio ha fornito un confronto dettagliato tra undici architetture CNN per la classificazione di immagini radiografiche del polso. L’uso della data augmentation ha significativamente migliorato le prestazioni. Inoltre, le mappe di attivazione si sono rivelate utili per comprendere il comportamento dei modelli. L’integrazione di ulteriori strategie di ottimizzazione e interpretabilità potrebbe migliorare ulteriormente l’affidabilità di queste tecniche in ambito clinico.