

Capitolul 1. Îmbunătățirea modelelor de diagnosticare imagistică

1.1 Introducere

În contextul avansării rapide a tehnologiilor bazate pe inteligența artificială, diagnosticarea imagistică a devenit una dintre cele mai promițătoare aplicații ale acestor inovații. Abordările tradiționale de diagnosticare, care depind adesea de expertiza umană și de interpretarea subiectivă a imaginilor medicale, pot fi îmbunătățite considerabil prin implementarea soluțiilor bazate pe AI. Acest capitol își propune să exploreze modul în care îmbunătățirea modelelor AI destinate diagnosticării imagistice poate fi realizată prin optimizarea seturilor de date, cu un accent deosebit pe tehnicile de preprocesare. Vom discuta despre importanța curățării datelor, eliminarea zgomotului, aplicarea diferitelor filtre vizuale și utilizarea augmentării datelor pentru a crește robustețea modelelor.

Dezvoltarea unui model AI eficient în diagnosticul medical nu se limitează doar la alegerea unei arhitecturi performante; calitatea datelor de intrare este esențială pentru rezultatele finale. Astfel, strategiile de preprocesare pot influența semnificativ acuratețea și fiabilitatea învățării automate. Prin urmare, este vital să investigăm nu doar seturile de date utilizate pe scară largă, cum ar fi ChestX-ray14, ISIC Archive și LIDC-IDRI, ci și metodologiile care transformă aceste date brute în input-uri optimizate pentru algoritmi de învățare profundă.

În plus, vom analiza cum aceste tehnici de îmbunătățire a datelor contribuie la o evaluare corectă a modelului AI, subliniind măsurile de performanță, cum ar fi AUC-ROC, sensibilitatea și specificitatea, care sunt imperativ de respectat pentru a asigura un diagnostic corect. Prin această abordare, capitolul va aduce o perspectivă cuprinzătoare asupra impactului atribuțiilor de preprocesare asupra eficienței modelelor AI în diagnosticarea imagistică, oferind nu doar o fundamentare teoretică, ci și o bază solidă pentru implementări practice în domeniul medical.

1.2 Îmbunătățirea Diagnosticului cu ChestX-ray14 și ResNet50

1.2.1 Setul de date ChestX-ray14 și utilizarea acestuia cu ResNet50

ChestX-ray14 este un set de date extins care cuprinde 112.120 imagini de radiografii toracice, provenite de la 30.805 pacienți unici, colectate între anii 1992 și 2015. Fiecare imagine este etichetată cu una dintre cele 14 patologii comune, inclusiv

pneumonie, emfizem și cancer pulmonar, fiind un instrument valoros în cercetarea în domeniul diagnosticării imagistice. Acest set de date este utilizat pentru antrenarea și evaluarea performanței modelelor de învățare profundă, cum ar fi ResNet50, un model de rețea neuronală convoluțională dezvoltat de Microsoft care a revoluționat domeniul clasificării imaginilor.

În cadrul utilizării ChestX-ray14 cu ResNet50, modelul este adaptat printr-un proces de fine-tuning, care implică ajustarea greutateilor și optimizarea rețelei prezente pe ImageNet pentru a se potrivi cerințelor specifice ale datelor de radiografie. Acest proces include modificarea ultimei straturi dens de ieșire a modelului pentru a corespunde numărului de patologii din setul de date (14 clase diferite). Fine-tuning permite modelului să învețe din caracteristicile particulare ale radiografiilor și să îmbunătățească acuratețea diagnosticării imaginilor, demonstrând o performanță semnificativ mai bună comparativ cu modele antrenate de la zero (2019).

1.3 Model Matematic și Justificare Teoretică

1.3.1 Arhitectura Matematică a Modelului

În contextul învățării profunde, dezvoltarea unui model matematic robust este esențială pentru obținerea unor rezultate precise și relevante. Unul dintre primii pași în definirea acestui model este stabilirea funcției obiectiv, care în esență reprezintă criteriul de optimizare utilizat în antrenarea modelului. În cazul clasificăției binare, o alegere frecventă este funcția de pierdere *Binary Cross-Entropy*, definită matematic ca:

$$L(\hat{y}, y) = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)]$$

unde y reprezintă eticheta reală, \hat{y} reprezintă predicția modelului, iar N este numărul total de observații.

Arhitectura rețelelor neuronale convoluționale (CNN) și mecanismul de auto-atenție implementat în Vision Transformers (ViT) sunt descrise printr-o serie de operații matematice specifice. Straturile convoluționale pot fi exprimate prin următoarea formulă:

$$A[i, j, k] = \sum_m \sum_n \sum_c X[i + m, j + n, c] \cdot K[m, n, c, k] + b[k]$$

unde:

- A este matricea de activare generată de stratul convoluțional.
- X reprezintă tensorul de intrare.
- K este kernelul convoluțional.

- b este biasul asociat.

Operația de convoluție (*) presupune aplicarea kernelului la fiecare porțiune a tensorului de intrare, ceea ce permite extragerea caracteristicilor esențiale ale imaginii. Aceasta este fundamentul procesării în rețelele CNN, oferind capacitatea de a detecta tipare și structuri specifice în datele de intrare.

În Vision Transformers (ViT), mecanismul de auto-atenție utilizează reprezentări matriciale și calculează similaritatea între token-uri pentru a determina relevanța fiecăruia. Formula mecanismului de auto-atenție este definită astfel:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

unde:

- Q (queries), K (keys) și V (values) reprezintă matrici ce conțin informații despre fiecare token din imagine.
- d_k este dimensiunea vectorilor de keys, utilizată pentru scalare, prevenind valori excesiv de mari care ar putea destabiliza funcția softmax.

Această funcție permite modelului să aloce greutate diferite fiecărui token în funcție de relevanța sa pentru sarcina dată, permițând astfel captarea relațiilor globale dintre părți ale imaginii.

1.3.2 Analiză Teoretică

Analizarea teoretică a arhitecturii modelului este esențială pentru justificarea alegerii sale în contextul sarcinilor pe care urmează să le îndeplinească. Capacitatea de generalizare este un concept crucial care se referă la abilitatea modelului de a face predicții corecte pe datele pe care nu le-a întâlnit anterior. O arhitectură bine proiectată va avea o capacitate mare de generalizare, evitând astfel supra-învățarea (over-fitting).

O metodă eficientă pentru reducerea over-fittingului constă în utilizarea tehnicilor de regularizare, care contribuie la creșterea robusteții modelului. Regularizarea L2, cunoscută și sub denumirea de *ridge regularization*, implică adăugarea unei penalizări pentru greutățile mari în funcția de pierdere. Aceasta poate fi exprimată matematic astfel:

$$L_{\text{total}} = L_{\text{original}} + \lambda \sum_i w_i^2$$

unde:

- L_{total} reprezintă funcția de pierdere totală după aplicarea regularizării.
- L_{original} este funcția de pierdere fără regularizare.
- λ este un hiper-parametru care controlează intensitatea regularizării.
- w_i sunt greutatele modelului.

Această tehnică penalizează valorile mari ale greutăților, încurajând modelul să generalizeze mai bine și să evite supra-învățarea pe datele de antrenament.

De asemenea, tehnica Dropout constă în dezactivarea aleatorie a anumitor neuroni în timpul fazei de antrenare, ceea ce forțează rețeaua să învețe reprezentări redundante și să devină mai tolerantă la variațiile în datele de intrare. Aceasta se poate exprima printr-o abordare probabilistică, permițând modelului să utilizeze o submulțime a neuronilor activi la fiecare iterație.

Modelul matematic care stă la baza arhitecturii selectate combină funcții matematice pentru definirea pierderilor, structuri complexe de rețea neuronală, și tehnici de regularizare, totul având scopul de a asigura o performanță optimă și o capacitate de generalizare eficientă pe datele necunoscute. Această fundamentare teoretică nu doar că sprijină deciziile de design ale modelului, ci și asigură aplicabilitatea sa în diverse aplicații practice în domeniul imagisticii medicale.