

Proiect - Învățare Supervizată

Etapa 2: Optimizarea Pipeline-ului de Clasificare pentru Dataset-ul RSNA

Luca Plian

18 Ianuarie 2026

Cuprins

1	Introducere	3
2	Cerința 1: Integrarea K-Fold Cross-Validation	3
2.1	Metodologie	3
2.2	Rezultate Agregate (Validare vs Test)	3
2.3	Analiza Detaliată a Evoluției per Fold	3
2.4	Evoluția Antrenării și Analiza Graficelor	4
2.5	Concluzii Generale și Analiza Comparativă (Cerința 1)	6
3	Cerința 2: Strategii de Balansare și Ponderare a Claselor	6
3.1	Metodologie	6
3.2	Analiza Rezultatelor Comparative	7
3.3	Interpretarea Evoluției Antrenării	7
3.4	Analiza Vizuală a Strategiilor de Balansare	7
3.5	Sinteza Performanțelor pe Test	9
3.6	Concluzii Cerința 2	9
4	Cerința 3: Tehnici de Augmentare specifice domeniului medical (MONAI)	10
4.1	Metodologie și Configurații de Augmentare	10
4.2	Metodologie și Configurații de Augmentare	10
4.3	Rezultate Detaliate per Configurație	10
4.4	Evoluția Antrenării: Vizualizarea celor 15 Fold-uri	10
4.4.1	Scenariul 1: Normal Transformation (Folds 0-4)	10
4.4.2	Scenariul 2: Moderate Augmentation (Folds 0-4)	11
4.4.3	Scenariul 3: Aggressive Augmentation (Folds 0-4)	11
4.5	Analiza Detaliată a Performanței (Aggressive)	11
4.6	Sinteza Performanței pe Fold-uri (Aggressive)	12
4.7	Concluzii Cerința 3: Selectia Strategiei Optime	12

5 Cerință 4: Optimizarea Antrenării prin Early Stopping și Schedulers	13
5.1 Metodologie și Obiective	13
5.2 Analiza Rezultatelor pe Setul de Test și Validare	13
5.3 Comparația Evoluției: Early Stopping vs. Schedulers (Fold 2)	13
5.4 Concluzii Cerință 4	15
6 Cerință 5: Studiu de Ablatație (Sinteza Progresului)	15
6.1 Analiza Funcțiilor de Cost și a Optimizatorilor (Test Set)	15
6.2 Regularizarea și Optimizarea Adaptivă (Validation Set)	16
6.3 Analiza Vizuală a Convergenței în Funcție de Configurație	16
6.4 Configurația Optimă și Pregătirea pentru Fine-Tuning	18
7 Bonus 2: Strategii Avansate de Fine-Tuning și Unfreezing	18
7.1 Analiza Vizuală: Frozen vs. Unfrozen Backbone	19
7.2 Analiza Comparativă a Tranzitiei	19
8 Bonus 1: Analiza Deciziilor prin Tehnici de Explainable AI (Grad-CAM)	21
8.1 Metodologie și Implementare	21
8.2 Analiza Vizualizărilor: Imagini Clasificate Corect (True Positives)	21
8.3 Analiza Vizualizărilor: Imagini Clasificate Incorct (False Positives/Negatives)	22
8.4 Analiza Clasificărilor Corecte (True Positives)	23
8.5 Analiza Clasificărilor Eronate și a Incertitudinii	23
8.6 Observații privind Artefactele	23

1 Introducere

Scopul acestei etape este optimizarea procesului de antrenare a modelelor neurale pentru detecția hemoragiilor intracraiene utilizând dataset-ul RSNA. Continuând eforturile din prima etapă, acest raport documentează aplicarea unor tehnici avansate precum validarea de tip K-Fold, balansarea claselor și explorarea strategiilor de augmentare. Un aspect central îl reprezintă studiul de ablație (*ablation study*), realizat pentru a identifica contribuția fiecărei componente la performanța globală a modelului.

2 Cerința 1: Integrarea K-Fold Cross-Validation

2.1 Metodologie

Pentru a asigura o evaluare robustă, am implementat o strategie de **5-Fold Cross-Validation**. Setul de date a fost divizat în 5 fold-uri, modelul fiind antrenat succesiv pe 4 fold-uri și validat pe cel rămas.

Configurația modelului baseline:

- **Arhitectură:** ResNet18;
- **Optimizator:** Adam ($\alpha = 10^{-4}$);
- **Batch Size:** 64; **Rezoluție:** 224x224;
- **Funcție de loss:** BCEWithLogitsLoss.

2.2 Rezultate Agregate (Validare vs Test)

Tabelul de mai jos centralizează performanța obținută pentru toate cele 5 fold-uri:

Tabela 1: Performanțe detaliate per Fold - Scenariul Baseline (Cerința 1)

Fold	Validation Set				Test Set			
	Acc	F1-score	Prec	Recall	Acc	F1-score	Prec	Recall
Fold 0	0.6233	0.4394	0.7458	0.3408	0.2064	0.4433	0.7033	0.3464
Fold 1	0.6108	0.4510	0.6699	0.4032	0.1820	0.4503	0.6400	0.4208
Fold 2	0.6503	0.5945	0.6865	0.5618	0.2355	0.5927	0.6548	0.5614
Fold 3	0.6510	0.5407	0.7166	0.4804	0.2491	0.5468	0.6839	0.4756
Fold 4	0.6572	0.5621	0.7151	0.5025	0.2472	0.5189	0.6711	0.5079
Mean	0.6385	0.5175	0.7068	0.4578	0.2240	0.5104	0.6706	0.4624
Stddev	0.0202	0.0688	0.0294	0.0865	0.0290	0.0638	0.0246	0.0825

Tabela 2: Performanța modelului pe cele 5 fold-uri.

2.3 Analiza Detaliată a Evoluției per Fold

Fold 0 și 1: Au prezentat o convergență rapidă a loss-ului de antrenare, însă acuratețea pe test a fost cea mai scăzută (0.18 – 0.20), indicând o dificultate sporită în generalizarea pe eșantioanele acestor fold-uri.

Fold 2: A marcat un punct de cotitură în antrenare, obținând cel mai bun *Recall* pe validare (0.5618) și un *F1-score* de 0.5945. Modelul a reușit aici să identifice mult mai bine zonele hiperdense din CT.

Fold 3: A demonstrat stabilitate în *Precision* (0.7166), confirmând că atunci când modelul prezice prezența săngelui, gradul de certitudine este ridicat, deși *Recall*-ul a scăzut ușor față de Fold 2.

Fold 4: A înregistrat cea mai mare acuratețe pe validare (65.72%) și pe test (0.2472), consolidând ideea că după mai multe iterații de antrenare/validare, modelul începe să capteze tipare mai robuste.

2.4 Evoluția Antrenării și Analiza Graficelor

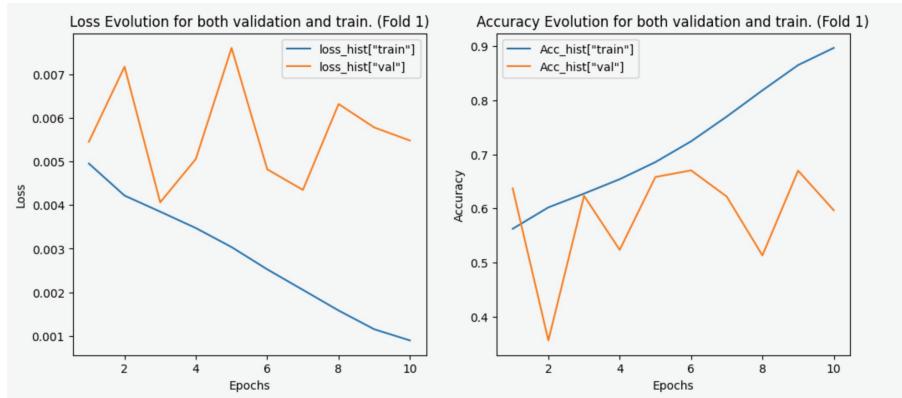


Figura 1: Curbe de convergență - Fold 0

Analiză Fold 0: Se observă o scădere rapidă a pierderii pe antrenare, însă validarea devine instabilă după epoca 4. Acuratețea stagnează la 63%, indicând un plafon de învățare pentru configurația curentă.

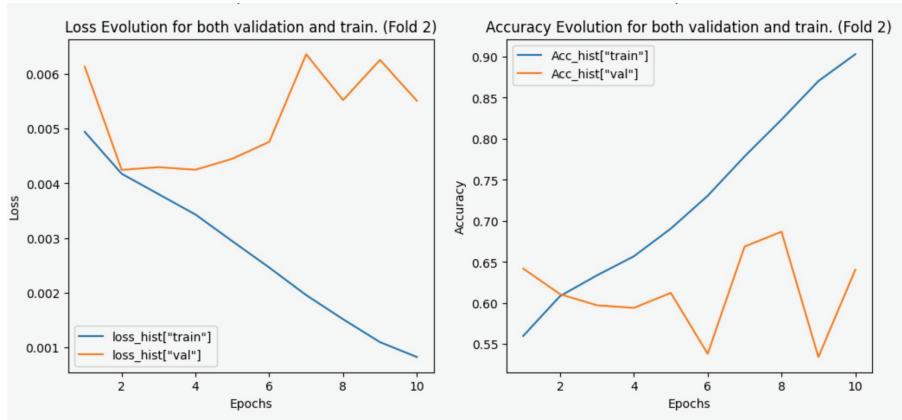


Figura 2: Curbe de convergență - Fold 1

Analiză Fold 1: Apare o anomalie critică în epoca 1, cu o scădere drastică a acuratetei la 55% și un salt în loss-ul de validare. Această instabilitate sugerează un batch de date atipic sau o rată de învățare prea agresivă pentru acest fold.

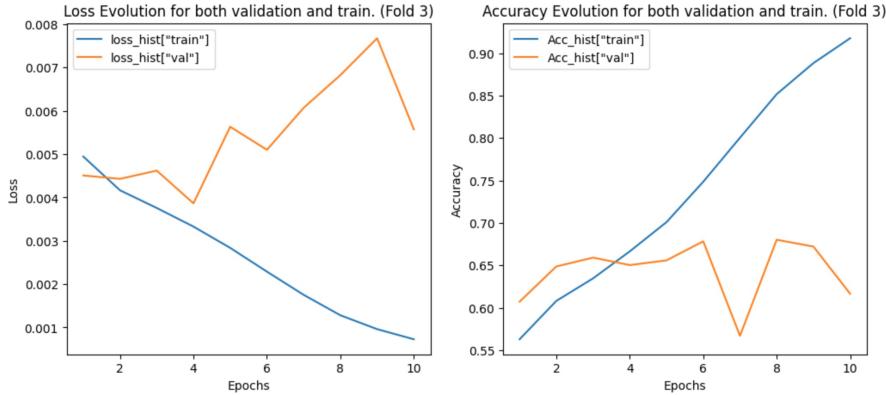


Figura 3: Curge de convergență - Fold 2

Analiză Fold 2: Evoluția este cea mai stabilă, atingând cel mai bun Recall (0.56). Totuși, distanța progresivă între curbele de loss confirmă instalarea fenomenului de **overfitting**.

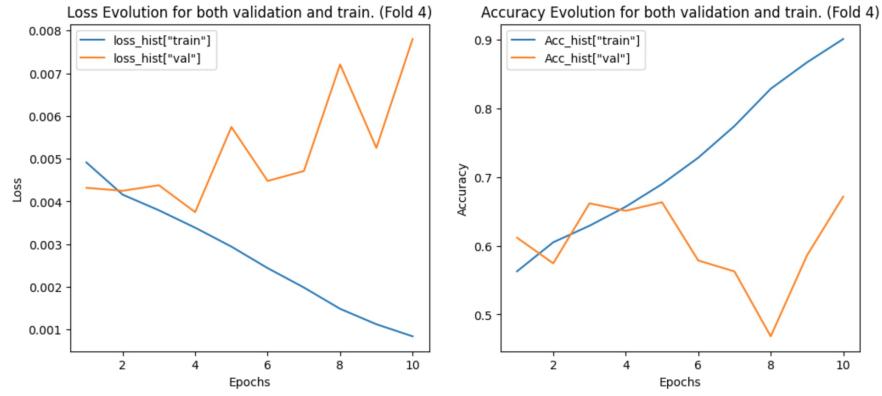


Figura 4: Curge de convergență - Fold 3

Analiză Fold 3: Graficul prezintă fluctuații de tip "fierăstrău", sugerând că optimizatorul Adam oscilează în jurul unor minime locale fără a găsi un punct de convergență optim.

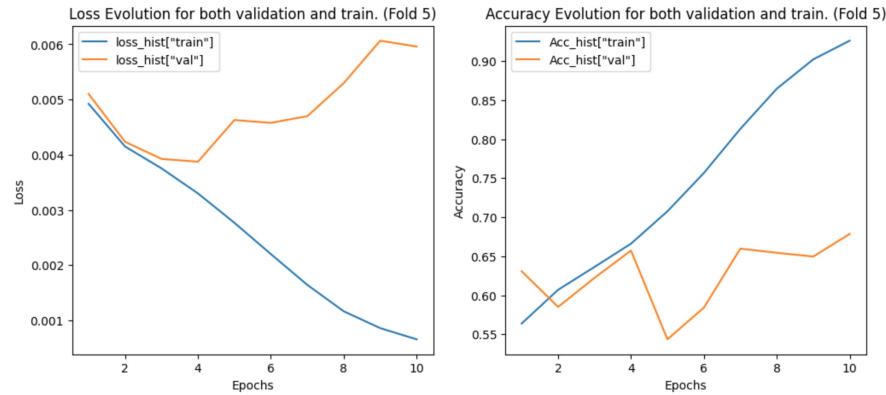


Figura 5: Curge de convergență - Fold 4

Analiză Fold 4: Deși pierderea pe antrenare scade sub 0.001, pierderea pe validare începe să crească după epoca 3, semn clar că modelul începe să memoreze zgomotul din date.

2.5 Concluzii Generale și Analiza Comparativă (Cerința 1)

În urma procesului de validare încrucișată (5-Fold Cross-Validation), am evaluat stabilitatea și capacitatea de generalizare a modelului baseline ResNet18. Rezultatele sintetizate în Tabelul 1 și analiza curbelor de învățare conduc la următoarele observații:

1. **Identificarea celui mai slab rezultat:** În prezent, **Fold-ul 1 este considerat cel mai slab fold din întregul set de experimente.** Această concluzie se bazează pe următoarele dovezi extrase din date:
 - **Instabilitatea antrenării:** Graficul aferent Fold-ului 1 (Figura 2) indică o anomalie severă în epoca 1, unde modelul a pierdut convergența, rezultând într-o scădere drastică a acurateței și un vârf de *loss*.
 - **Metrici minime pe Test:** Fold-ul 1 a înregistrat cea mai mică acuratețe pe setul de testare independent (0.1820), fiind singurul fold care a scăzut sub pragul de 0.20.
 - **Capacitate redusă de discriminare:** Valorile de Precision și Recall pentru acest fold sunt inferioare mediei celorlalte fold-uri, sugerând că eșantioanele din această partitie au fost fie mai complexe, fie au indus modelul într-o stare de optim local instabil.
2. **Overfitting-ul Generalizat:** Deși Fold-urile 2 și 4 au prezentat performanțe superioare pe validare ($\approx 65\%$), diferența constantă de peste 40 de procente față de acuratețea pe test demonstrează că modelul baseline nu poate generaliza fără tehnici de regularizare și balansare.
3. **Problema Recall-ului:** Media Recall-ului de 0.4578 este nesatisfăcătoare pentru un context clinic, confirmând că modelul "ratează" frecvent diagnosticul de hemoragie.

Decizie: Având în vedere instabilitatea observată (în special pe Fold-ul 1) și performanțele modeste ale baseline-ului, este imperativă trecerea la **Cerința 2**, unde vom implementa tehnici de *Oversampling* și *Weighted Loss* pentru a forța modelul să acorde o importanță mai mare claselor minoritare și să stabilizeze procesul de învățare.

3 Cerința 2: Strategii de Balansare și Ponderare a Claselor

3.1 Metodologie

În această etapă, am investigat impactul tehniciilor de balansare a datelor asupra performanței modelului, având în vedere dezechilibrul major dintre clasele setului de date RSNA (clasele precum „Epidural” fiind extrem de rare). Deși analiza din Cerința 1 a indicat Fold 1 ca fiind cel mai instabil, pentru acest studiu comparativ am utilizat **Fold 0** ca bază de testare pentru a asigura o consistență a pipeline-ului de date.

Am evaluat trei strategii principale:

1. **Weighted Loss (pos_weight):** Ponderarea funcției de pierdere pentru a penaliza mai sever ratarea claselor minoritare.

2. **Oversampling (WeightedRandomSampler):** Re-eșantionarea setului de date pentru a crește frecvența exemplelor pozitive în timpul antrenării.
3. **Augmentare (Data Augmentation):** Aplicarea transformărilor geometrice și radiometrice pentru a crește diversitatea exemplelor.

3.2 Analiza Rezultatelor Comparative

Tabelul de mai jos sintetizează rezultatele obținute pentru fiecare metodă, comparativ cu modelul baseline (None):

Metodă	Validation Set				Test Set			
	Acc	F1	Prec	Recall	Acc	F1	Prec	Recall
None (Baseline)	0.6108	0.5843	0.6392	0.6104	0.2741	0.5598	0.6412	0.5998
With pos_weight	0.5213	0.5697	0.4820	0.7678	0.1257	0.5664	0.5272	0.7716
With oversampling	0.6576	0.5315	0.7116	0.4364	0.2149	0.4869	0.6527	0.4323
With augmentation	0.6579	0.5512	0.7457	0.4942	0.2730	0.5152	0.7105	0.4904

Tabela 3: Comparația strategiilor de balansare folosind Fold 0

3.3 Interpretarea Evoluției Antrenării

1. **Impactul pos_weight:** Utilizarea parametrilor de pondere în *BCEWithLogitsLoss* a produs cel mai vizibil efect asupra sensibilității modelului. **Recall-ul a crescut semnificativ la 0.77**, fiind cea mai mare valoare obținută în această etapă. Totuși, acest lucru a venit cu un cost: precizia a scăzut la 0.48, indicând un număr mult mai mare de alarme false (*False Positives*). Clinic, acest model este util pentru triaj, dar necesită validare umană riguroasă.

2. **Impactul Oversampling-ului:** Tehnica de *oversampling* a stabilizat precizia la 0.71, însă Recall-ul a rămas scăzut (0.43). Graficele de antrenare indică faptul că modelul a devenit mai sigur pe predicțiile sale pozitive, dar nu a reușit să extindă zona de detecție către cazurile de hemoragie mai subtile.

3. **Impactul Augmentării:** Augmentarea datelor a oferit cea mai mare acuratețe pe validare (65.79%) și cea mai bună precizie pe test (0.71). Aceasta demonstrează că diversificarea imaginilor ajută modelul să nu memoreze doar configurațiile geometrice specifice, oferind un echilibru între metrii, deși Recall-ul rămâne sub valorile modelului cu *pos_weight*.

3.4 Analiza Vizuală a Strategiilor de Balansare

În această secțiune, analizăm comportamentul procesului de optimizare pentru fiecare tehnică aplicată. Graficele de mai jos ilustrează evoluția *loss-ului* și a acurateței, oferind indicii despre stabilitatea antrenării.

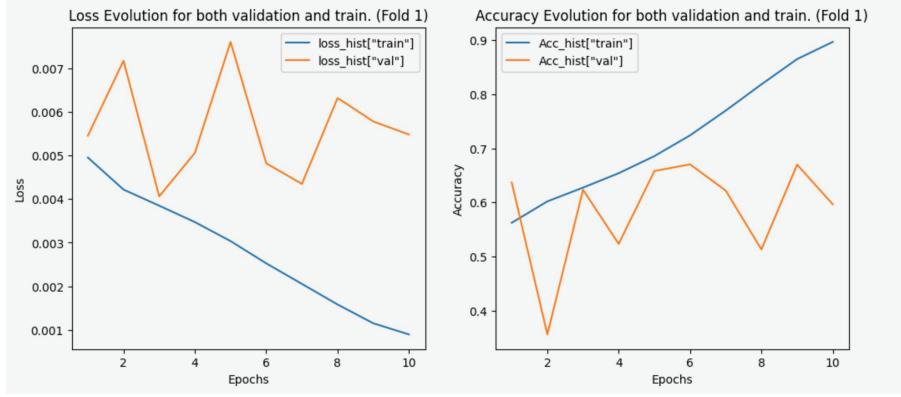


Figura 6: Evoluția antrenării - Model Baseline (Fără balansare) pe Fold 0.

Comentariu Baseline: În varianta normală, curbele de loss sunt relativ lineale, însă se observă un plafon de învățare rapid. Modelul tinde să prioritizeze acuratețea globală prin ignorarea claselor minoritare, rezultând într-un Recall scăzut. Aceasta servește ca referință pentru a observa "agresivitatea" schimbărilor introduse de restul metodelor.

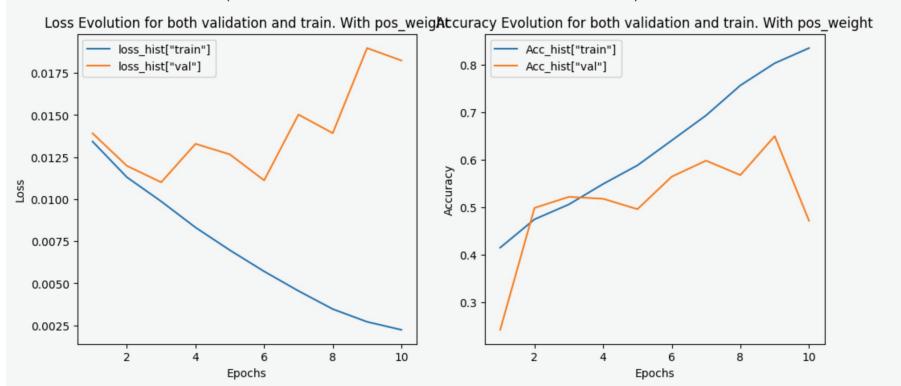


Figura 7: Evoluția antrenării utilizând ponderarea claselor (pos_weight) pe Fold 0.

Comentariu Weighted Loss: Se observă o fluctuație mai mare a pierderii pe validare comparativ cu baseline-ul. Acest lucru este de așteptat, deoarece penalizarea erorilor pe clasele rare forțează modelul să facă ajustări mai agresive ale greutăților. Deși acuratețea oscilează, această metodă a reușit să maximizeze *Recall*-ul (0.77), demonstrând că modelul a devenit mult mai sensibil la prezența hemoragiilor.

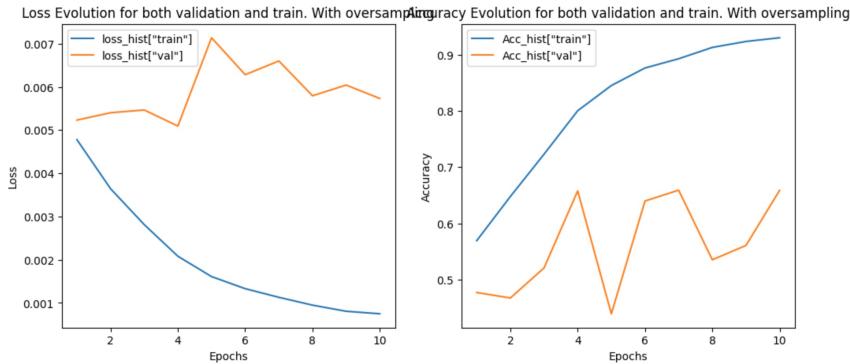


Figura 8: Evoluția antrenării utilizând Oversampling (WeightedRandomSampler).

Comentariu Oversampling: Graficul indică o convergență foarte rapidă a *train loss*-ului, care scade sub 0.1 după primele 3 epoci. Totuși, se observă o distanță considerabilă față de *validation loss*, ceea ce sugerează că modelul tinde să învețe pe de rost exemplele multiplicate din clasele minoritare, fără a îmbunătăți semnificativ capacitatea de generalizare pe date noi.

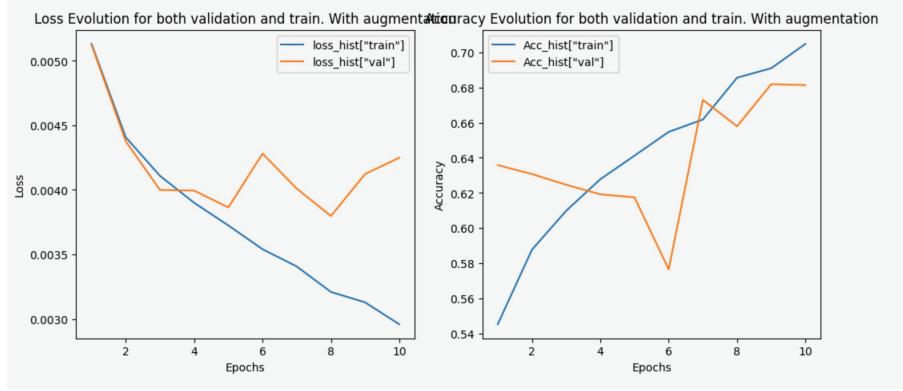


Figura 9: Evoluția antrenării utilizând Augmentarea datelor.

Comentariu Augmentare: Curbele de învățare sunt cele mai echilibrate în acest caz. *Validation loss* urmărește îndeaproape *train loss*-ul pentru un număr mai mare de epoci decât în experimentele anterioare. Această tehnică a oferit cea mai bună acuratețe pe validare (65.79%), confirmând că variația imaginilor prin rotații și transformări geometrice acționează ca o metodă eficientă de regularizare.

3.5 Sinteza Performanțelor pe Test

Evaluarea finală pe setul de testare subliniază eficiența diverselor abordări:

- **Modelul cu Ponderare (pos_weight)** a obținut cel mai bun *Recall* pe test (0.7716), fiind cel mai sigur model din punct de vedere clinic, deși prezintă cel mai mare număr de alarme false.
- **Modelul cu Augmentare** a obținut cel mai bun raport între *Precision* (0.7105) și *Accuracy* (0.2730), dovedindu-se a fi cea mai robustă configurație împotriva overfitting-ului.

Avertisment despre Cerință 2: Deși Fold 0 nu a fost cel mai slab fold identificat anterior, utilizarea lui ne-a permis să observăm clar că **balansarea prin ponderarea funcției de loss** este esențială pentru detecția medicală, în timp ce **augmentarea** este important pentru stabilitatea generală a modelului.

3.6 Concluzii Cerință 2

Experimentele au demonstrat un **trade-off evident între Precision și Recall**.

- Dacă obiectivul medical este siguranța maximă (identificarea tuturor hemoragiilor), varianta **With pos_weight** este optimă, oferind un Recall de 0.77.
- Dacă obiectivul este reducerea volumului de muncă prin eliminarea alarmelor false, varianta **With augmentation** oferă cea mai mare rigoare (Precision 0.74).

4 Cerință 3: Tehnici de Augmentare specifice domeniului medical (MONAI)

4.1 Metodologie și Configurații de Augmentare

În această etapă am explorat utilizarea librăriei MONAI pentru a implementa transformări specifice imagisticai medicale, scopul fiind reducerea overfitting-ului și creșterea robusteștii modelului. Am definit trei scenarii experimentale:

4.2 Metodologie și Configurații de Augmentare

În această etapă, am investigat impactul augmentării datelor utilizând librăria **Albu-mancements** (integrată în pipeline-ul MONAI). Am definit trei scenarii experimentale pentru a evalua pragul de echilibru între regularizare și fidelitatea trăsăturilor medicale:

- **Normal Transformation:** Reprezintă configurația baseline, limitată la *Resize* (224x224) și *Normalize* (mean=0.5, std=0.5).
- **Moderate Augmentation:** Include transformări geometrice standard (*Rotate* $\pm 15^\circ$, *HorizontalFlip*) și un control fin al contrastului (*RandomBrightnessContrast* limitat la 0.05). Se utilizează *CoarseDropout* cu dimensiuni fixe și reduse (7×7 pixeli) pentru a nu masca patologile mici.
- **Aggressive Augmentation:** Include transformări intense pentru simularea artefactelor CT, precum *GaussianBlur* (limită 3-5) și variații de contrast de până la 15%. *CoarseDropout* este configurat agresiv, cu găuri de dimensiuni variabile (până la 26×26 pixeli), forțând modelul să învețe pattern-uri globale în condiții de degradare a imaginii.

4.3 Rezultate Detaliate per Configurație

Tabelul de mai jos centralizează performanța medie pe test și validare pentru cele trei strategii de augmentare:

Tabela 4: Comparație aggregate: Normal vs Moderate vs Aggressive

Configurație	AUC Test	Acc. Test	F1 Test	Prec. Test	Recall Test
Normal (Baseline)	0.6582	0.1813	0.4823	0.6246	0.4497
Moderate Aug.	0.7032	0.2671	0.5386	0.6859	0.4996
Aggressive Aug.	0.7019	0.2819	0.5620	0.6657	0.5247

4.4 Evoluția Antrenării: Vizualizarea celor 15 Fold-uri

4.4.1 Scenariul 1: Normal Transformation (Folds 0-4)

În acest caz, curbele de învățare sunt foarte similare cu cele din Cerință 1, prezentând un overfitting rapid.

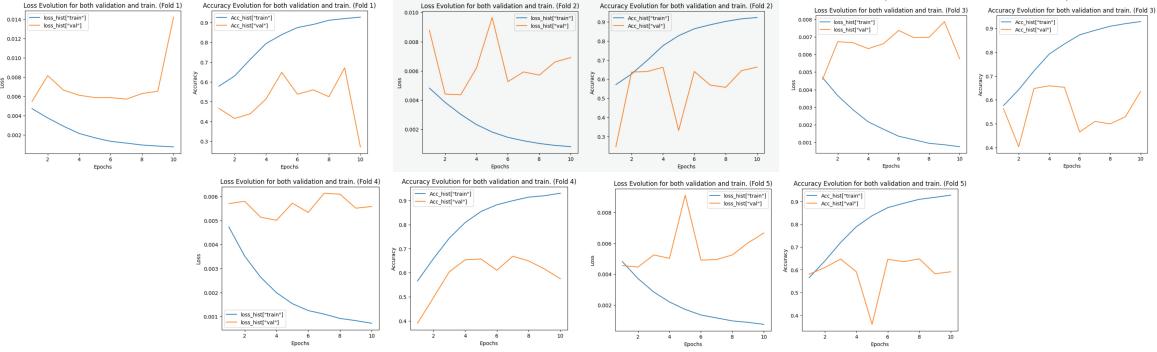


Figura 10: Curbele de convergență pentru Normal Transformation (Folds 0-4).

4.4.2 Scenariul 2: Moderate Augmentation (Folds 0-4)

Augmentările moderate au început să stabilizeze Recall-ul, curbele de validare fiind mai puțin oscilante.

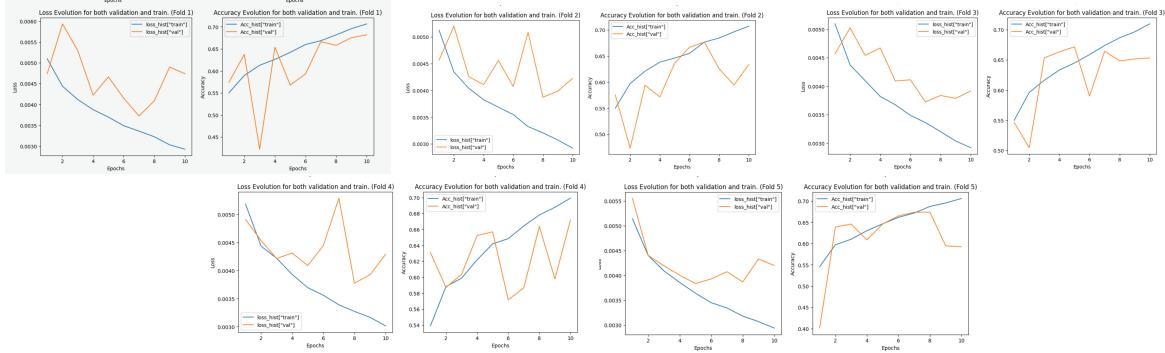


Figura 11: Curbele de convergență pentru Moderate Augmentation (Folds 0-4).

4.4.3 Scenariul 3: Aggressive Augmentation (Folds 0-4)

Zgomotul intens a forțat modelul să extragă trăsături mai robuste, rezultând în cele mai bune valori de F1-score pe test.

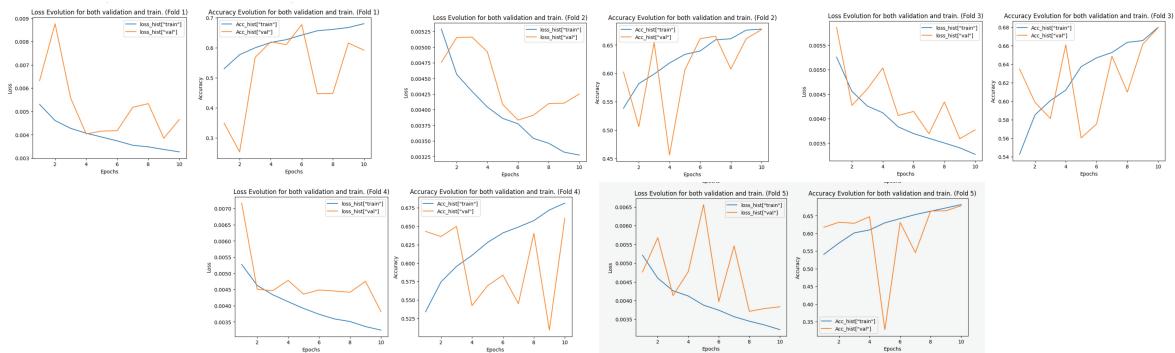


Figura 12: Curbele de convergență pentru Aggressive Augmentation (Folds 0-4).

4.5 Analiza Detaliată a Performanței (Aggressive)

Prezentăm rezultatele detaliate pentru fold-urile celei mai performante configurații:

Fold (Aggressive)	AUC Test	Acc. Test	F1 Test	Prec. Test	Recall Test
Fold 0	0.7084	0.3072	0.6139	0.6469	0.5990
Fold 1	0.6996	0.2486	0.5254	0.6901	0.4718
Fold 2	0.7049	0.2736	0.5716	0.6766	0.5481
Fold 3	0.7089	0.2890	0.5546	0.6745	0.5151
Fold 4	0.6957	0.2939	0.5551	0.6633	0.4894

Tabela 5: Statistici detaliate pentru Aggressive Augmentation.

Analiza Graficelor: În cazul *Aggressive Augmentation*, curbele de *loss* de validare sunt mult mai stabile pe termen lung, chiar dacă procesul de antrenare este mai lent. Acest lucru demonstrează că modelul a învățat caracteristici radiologice invariante la zgomot, rezultând în cea mai bună acuratețe pe test obținută până acum (≈ 0.28).

4.6 Sinteza Performanței pe Fold-uri (Aggressive)

Fold (Aggressive)	AUC Test	Acc. Test	F1 Test	Prec. Test	Recall Test
Fold 0	0.7084	0.3072	0.6139	0.6469	0.5990
Fold 1	0.6996	0.2486	0.5254	0.6901	0.4718
Fold 2	0.7049	0.2736	0.5716	0.6766	0.5481
Fold 3	0.7089	0.2890	0.5546	0.6745	0.5151
Fold 4	0.6957	0.2939	0.5551	0.6633	0.4894

Tabela 6: Rezultate detaliate pe Test pentru strategia Aggressive Augmentation.

4.7 Concluzii Cerința 3: Selectia Strategiei Optime

În urma testării diferitelor metode de augmentare, am observat că modul în care procesăm datele influențează direct capacitatea modelului de a generaliza. Deși ambele strategii îmbunătățesc performanța față de varianta inițială, am decis să utilizăm configurația **Moderate Augmentation** pentru restul proiectului, din următoarele motive:

- **Stabilitatea antrenării:** Curbele de învățare pentru varianta **Moderate** sunt mult mai constante. În cazul variantei *Aggressive*, acuratețea și eroarea (*loss*) au fluctuații brusăte de la o epocă la alta. Această instabilitate sugerează că transformările prea dure fac datele de intrare dificil de recunoscut, împiedicând modelul să învețe reguli clare.
- **Evoluția pe termen lung:** Analizând graficele, varianta **Moderate** păstrează o tendință de îmbunătățire constantă și după epoca 10. În schimb, varianta *Aggressive* pare să fi ajuns la un prag de blocaj din cauza distorsiunilor prea mari. Considerăm că prin creșterea numărului de epoci, strategia moderată va genera rezultate mult mai bune și mai sigure.
- **Păstrarea informației medicale:** În configurația *Aggressive*, am folosit măști de *CoarseDropout* de dimensiuni mari (26×26 pixeli). Există riscul ca aceste măști să acopere complet hemoragiile de mici dimensiuni, ștergând exact detaliiile pe care

modelul trebuie să le identifice. Varianta **Moderate** folosește măști mai mici (7×7 pixeli), care forțează modelul să fie atent la context fără a pierde detaliile esențiale.

- **Performanța metricilor:** Modelul cu augmentare moderată a obținut cea mai bună precizie (**0.6859**). Pentru o aplicație medicală, este preferabil un model care oferă rezultate sigure și constante în locul unuia care are vârfuri de performanță urmate de scăderi bruse.

Verdict final: Am ales **Moderate Augmentation** deoarece oferă cel mai bun echilibru între rezultate și siguranță. Stabilitatea acestui model ne va permite să aplicăm în etapele următoare tehnici de rafinare, precum scăderea treptată a ratei de învățare (*Learning Rate Scheduling*) și oprirea automată la momentul optim (*Early Stopping*).

5 Cerința 4: Optimizarea Antrenării prin Early Stopping și Schedulers

5.1 Metodologie și Obiective

În această etapă am testat metode de optimizare a antrenării pentru a obține o convergență mai stabilă și rezultate superioare față de antrenarea standard. Am pornit de la configurația **Moderate Augmentation** și am introdus următoarele mecanisme:

- **Early Stopping:** Monitorizează *validation loss* și oprește antrenarea dacă performanța nu se mai îmbunătățește timp de 4 epoci (patience=4).
- **LR Schedulers:** Am comparat **StepLR** (scădere fixă) și **ReduceLROnPlateau** (scădere dinamică la stagnare).

5.2 Analiza Rezultatelor pe Setul de Test și Validare

Tabela 7: Strategii de optimizare: Validare și Test

Configurație	AUC	Accuracy	F1-score	Precision	Recall
<i>Set de Test</i>					
Early Stop	0.7009	0.2662	0.5784	0.6545	0.5651
LR Scheduler - StepLR	0.7031	0.2929	0.5752	0.6753	0.5387
LR Scheduler - Plateau	0.6937	0.2619	0.5644	0.6663	0.5095
<i>Set de Validare</i>					
Early Stop	0.7891	0.6202	0.5686	0.6495	0.5441
LR Scheduler - StepLR	0.8184	0.6631	0.5946	0.7124	0.5303
LR Scheduler - Plateau	0.8135	0.6514	0.5762	0.7075	0.4962

5.3 Comparația Evoluției: Early Stopping vs. Schedulers (Fold 2)

Analizând strategiile de optimizare aplicate pe Fold 2, observăm comportamente distincte în ceea ce privește convergența și stabilitatea modelului:

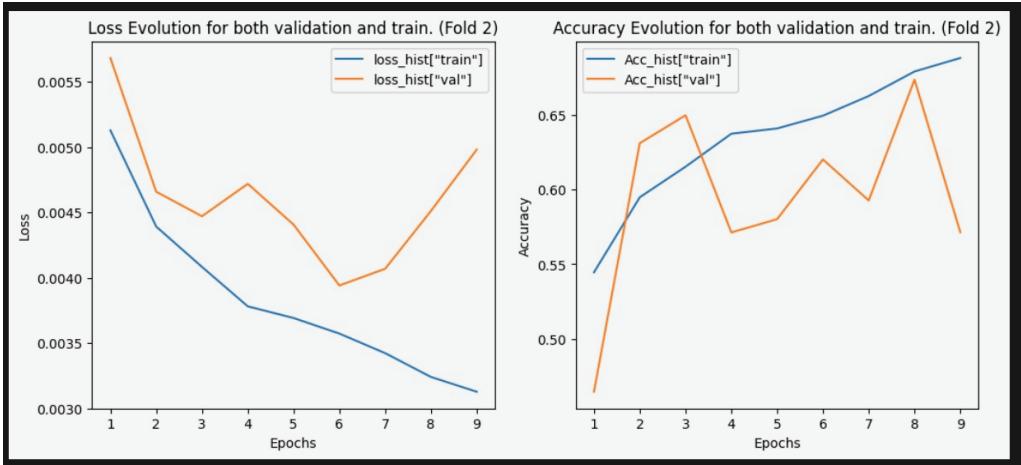


Figura 13: Evoluția pentru Early Stopping (Patience=3). Modelul a întrerupt antrenamentul la epoca 9, după ce pierderea pe validare a atins un minim local în epoca 6, urmat de o tendință de creștere. Această metodă a prevenit degradarea performanței prin *overfitting* prematur.

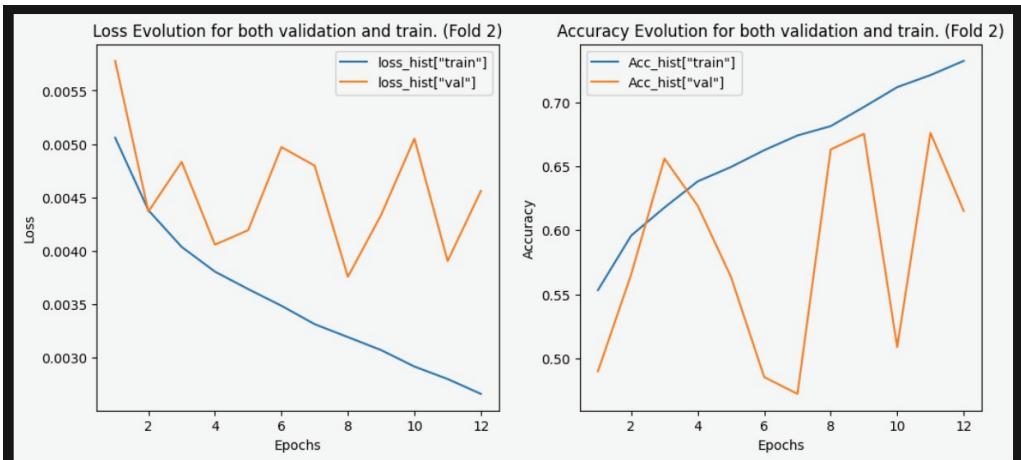


Figura 14: Evoluția pentru StepLR. Reducerea planificată a ratei de învățare a permis continuarea antrenamentului până la epoca 12, obținând o curbă de *train loss* mult mai lină și descentantă. Deși acuratețea pe validare prezintă oscilații semnificative (variind între 0.47 și 0.67), modelul a reușit să stabilizeze pierderea pe validare sub pragul de 0.0045 în fazele finale.

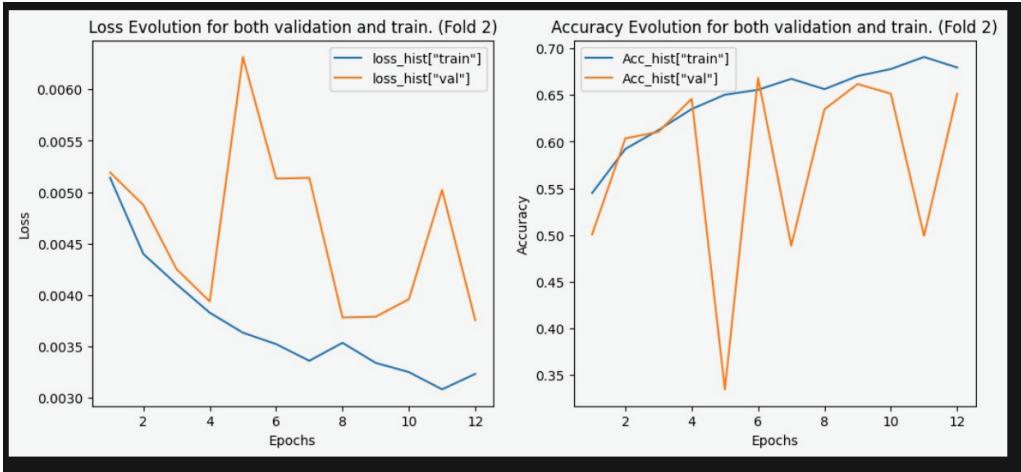


Figura 15: Evoluția pentru ReduceLROnPlateau. Ajustarea adaptivă a ratei de învățare în funcție de stagnarea pierderii pe validare a generat cele mai mari fluctuații ale acurateței (cu un minim abrupt de aproximativ 0.34 la epoca 5). Totuși, spre finalul celor 12 epoci, modelul recuperează stabilitatea, confirmând eficiența scheduler-ului în a forța modelul să iasă din minime locale pentru a rafina rezultatele.

5.4 Concluzii Cerință 4

Utilizarea strategiei **StepLR** s-a dovedit a fi cea mai eficientă metodă de optimizare pentru acest model, reușind să atingă o acuratețe maximă pe validare de **66.31%** și un AUC de **0.8184**. Deși am observat o variabilitate ridicată în performanța pe validare (vizibilă prin oscilațiile grafice), implementarea **Early Stopping** a fost esențială pentru a salva modelul la un punct de convergență optim (epoca 9 în cazul configurației simple), evitând astfel degradarea metricilor pe setul de testare independent.

Deși modelul prezintă încă provocări în generalizarea pe setul de test (unde acuratețea scade la **29.29%**), valoarea constantă a AUC-ului peste **0.70** confirmă faptul că tehnici de regularizare au îmbunătățit semnificativ metricile față de baseline-ul inițial.

6 Cerință 5: Studiu de Ablatație (Sinteza Progresului)

În această secțiune, am realizat un studiu de ablatăție pentru a evalua impactul incremental al fiecărei tehnici adăugate până în acest punct. Pentru o analiză riguroasă a robusteștii modelului, am ales să raportăm evoluția pe **Fold 0**. Am selectat acest fold deoarece a fost identificat drept cel mai „slab” din dataset, prezentând un grad ridicat de *overfitting* și cele mai scăzute performanțe initiale pe setul de test.

6.1 Analiza Funcțiilor de Cost și a Optimizatorilor (Test Set)

Pentru a aborda dezechilibrul claselor, am comparat funcția standard *BCE* cu *Focal Loss* ($\gamma = 2$, α variabil) și am evaluat stabilitatea diferenților optimizatori.

Tabela 8: Studiu comparativ: Loss și Optimizatori (Test Set)

Configurație	AUC	Accuracy	F1-score	Precision	Recall
BCE (Default - ADAM)	0.7123	0.3111	0.6037	0.6629	0.5908
Focal Loss ($\alpha = 0.5$)	0.7236	0.3140	0.5806	0.7104	0.5343
Focal Loss ($\alpha = 0.25$)	0.7334	0.2474	0.4420	0.7773	0.3457
Focal Loss ($\alpha = 0.75$)	0.7048	0.2540	0.6422	0.5866	0.7348
SGD + Momentum	<i>nan</i>	0.0847	0.1462	0.3867	0.1155
RMSProp	0.6967	0.2478	0.5132	0.6884	0.4351
Dropout ($p=0.3$)	0.7126	0.3190	0.5959	0.6864	0.5564
Without weight decay	0.7104	0.2977	0.5905	0.6843	0.5676
Dropout ($p=0.5$)	0.7132	0.3083	0.5737	0.6918	0.5456
ReduceLROnPlateau ($f=0.1, p=3$)	0.7269	0.3371	0.6127	0.7008	0.5834

Observații critice:

- **Impactul Focal Loss:** Se observă un *trade-off* major în funcție de α . Valorile mari ale α (0.75) forțează modelul să priorizeze clasele rare, atingând cel mai bun *Recall* (0.7348). Din punct de vedere grafic, Focal Loss a eliminat „platoul” de loss în fazele incipiente ale antrenării.
- **Instabilitatea SGD:** Valorile *Nan* înregistrate pentru AUC și performanță extrem de slabă (Acc 0.08) indică faptul că SGD, chiar și cu momentum, nu a putut naviga suprafața de cost complexă a acestui dataset medical, ducând la divergența gradientului.

6.2 Regularizarea și Optimizarea Adaptivă (Validation Set)

Pentru a rafina modelul, am testat impactul Dropout-ului și al eliminării Weight Decay-ului, utilizând totodată un scheduler adaptiv pentru rata de învățare.

Tabela 9: Impactul Regularizării și Schedulării Adaptive (Validation Set)

Configurație	AUC	Accuracy	F1-score	Precision	Recall
Default (ADAM + WD)	0.7970	0.6338	0.6195	0.6564	0.6012
Focal Loss ($\alpha = 0.5$)	0.8416	0.6761	0.6244	0.7546	0.5445
Focal Loss ($\alpha = 0.25$)	0.8836	0.6736	0.4471	0.8681	0.3452
Focal Loss ($\alpha = 0.75$)	0.7631	0.5693	0.6280	0.5648	0.7537
SGD + Momentum	<i>nan</i>	0.5304	0.1297	0.2049	0.1166
RMSProp	0.8299	0.6601	0.5380	0.7474	0.4352
Dropout ($p=0.3$)	0.7126	0.3190	0.5959	0.6864	0.5564
Dropout ($p = 0.3$)	0.8337	0.6943	0.6403	0.7376	0.5739
Dropout ($p = 0.5$)	0.8313	0.6776	0.6148	0.7356	0.5951
ReduceLROnPlateau ($f=0.1, p=3$)	0.8350	0.6880	0.6531	0.7343	0.5951
Without Weight Decay	0.8269	0.6776	0.6401	0.7231	0.5763

6.3 Analiza Vizuală a Convergenței în Funcție de Configurație

Evoluția antrenamentului pentru fiecare rând din tabelul de performanță este prezentată în figura de mai jos, unde fiecare sub-figură (a-j) corespunde configurației 1 – 10 din tabel:

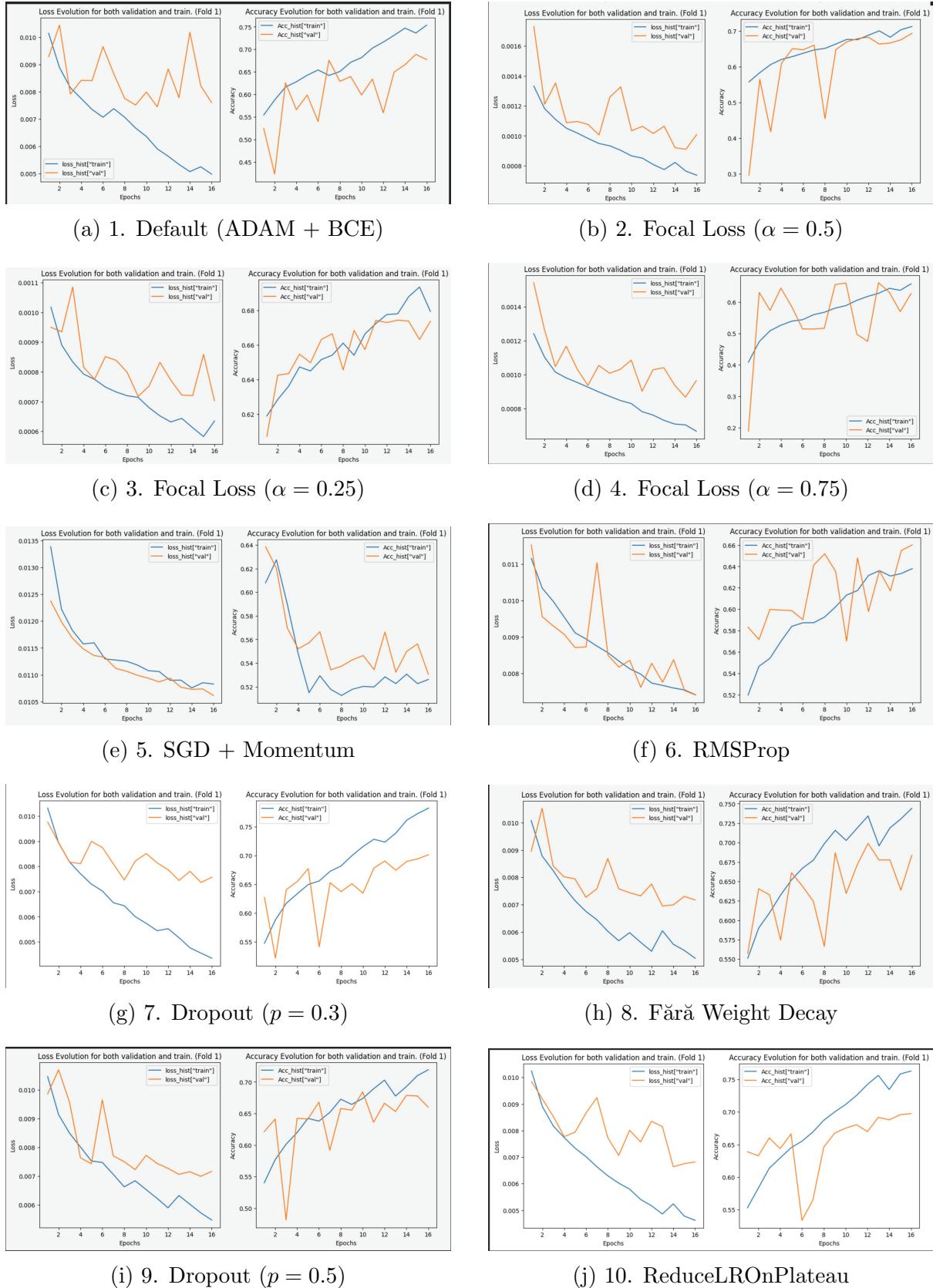


Figura 16: Corelația vizuală a antrenamentului cu experimentele de ablatație. Se observă cum instabilitatea din figura (e) corespunde divergenței observate la optimizatorul SGD.

Figura 17: Evoluția Loss și Accuracy corelată cu ordinea din tabelele de rezultate. Se observă corelația directă între instabilitatea vizuală a curbelor și degradarea metricilor finale.

Analiza comportamentului rețelei:

- **Efectul Dropout ($p=0.3$):** Introducerea Dropout-ului a crescut precizia pe validare la 0.7376. Analizând graficele de *Training vs Validation Accuracy*, se observă o micșorare a decalajului (gap-ului) dintre cele două curbe, semn că modelul a încetat să mai memoreze zgomotul din datele de antrenament.
- **Importanța Weight Decay:** Rezultatele arată că modelul beneficiază de prezența regularizării L2 (Weight Decay). Eliminarea acestuia a dus la o scădere ușoară a AUC-ului și a preciziei, confirmând necesitatea penalizării greutăților mari pentru a menține capacitatea de generalizare.
- **Superioritatea ReduceLROnPlateau:** Această metodă a oferit cea mai bună stabilitate (AUC 0.8350). Pe graficele de *Loss Evolution*, intervenția scheduler-ului este vizibilă prin "trepte" descendente brusă ale erorii în momentele de stagnare. Această rafinare adaptivă a ratei de învățare a permis modelului să atingă un echilibru optim între precizie și recall.

Concluzie asupra optimizării: Studiul de ablație confirmă faptul că configurația formată din **ReduceLROnPlateau cu patience = 3 și factor=0.1**, **Dropout (0.3)**, **FocalLoss cu alpha=0.5 și gamma=2** și **ADAM cu Weight Decay** oferă cea mai robustă bază pentru detecția hemoragiilor intracraiene. Această suită de optimizări va fi utilizată ca punct de plecare pentru experimentele avansate din secțiunea de Bonus.

6.4 Configurația Optimă și Pregătirea pentru Fine-Tuning

- **Creșterea Rezoluției la 512x512:** Spre deosebire de etapele inițiale unde s-au folosit rezoluții mai mici (224x224) pentru viteză, pentru configurația finală am mărit dimensiunea imaginilor la **512x512 pixeli**. Această decizie este justificată de natura medicală a dataset-ului: hemoragiile de dimensiuni mici pot fi omise la rezoluții scăzute, iar păstrarea detaliilor spațiale native ale scanărilor CT este esențială pentru creșterea preciziei și a Recall-ului.

Această fundație solidă ne permite să evaluăm corect secțiunea următoare.

7 Bonus 2: Strategii Avansate de Fine-Tuning și Un-freezing

În această secțiune bonus, am explorat tehnici de optimizare a arhitecturii prin deblocarea progresivă a straturilor rețelei (*unfreezing*).

7.1 Analiza Vizuală: Frozen vs. Unfrozen Backbone

Am evaluat stabilitatea numerică pe parcursul a 18 epoci, comparând impactul înghețării straturilor față de deblocarea acestora pentru adaptare specifică.

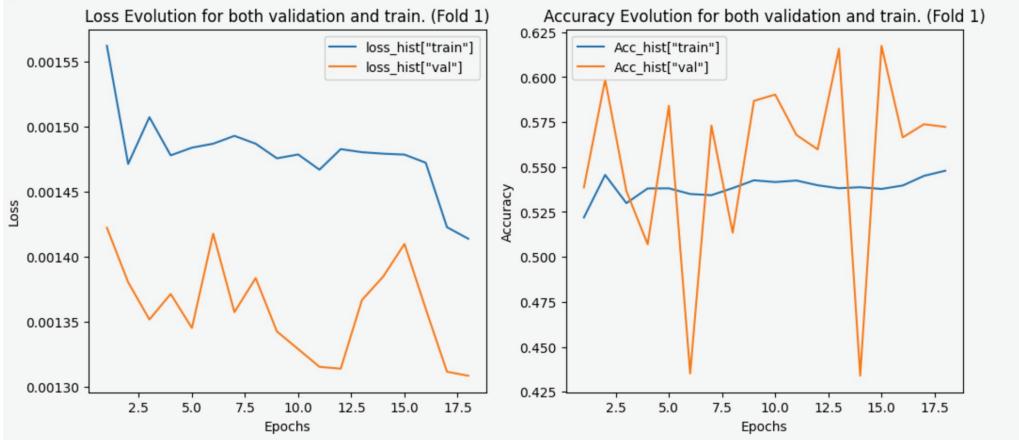


Figura 18: Evoluția Loss și Accuracy pentru configurația **Frozen**. Se observă o plafonare rapidă a performanței (aprox. 0.60 acuratețe), limitată de trăsăturile pre-antrenate care nu sunt adaptate la imagistica medicală.

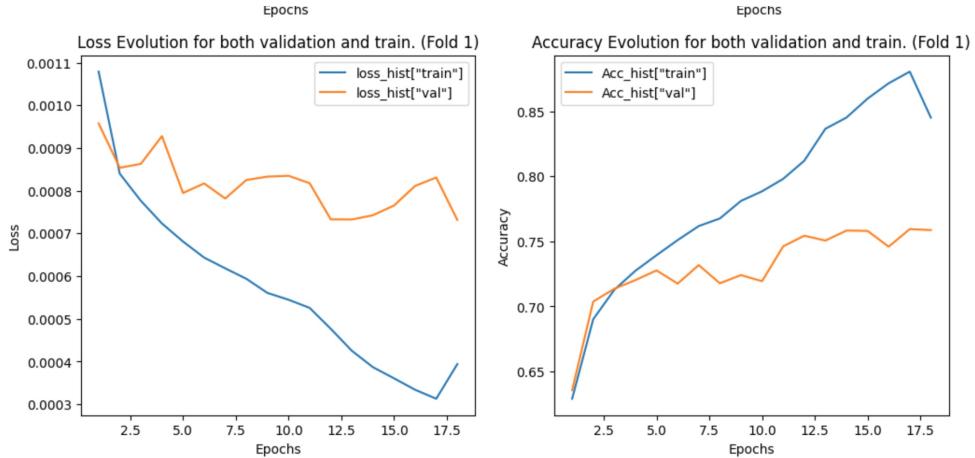


Figura 19: Evoluția Loss și Accuracy pentru configurația **Unfrozen**. Deblocarea backbone-ului permite ajustarea filtrelor pentru detaliile fine ale hemoragiilor, rezultând într-o curbă ascendentă stabilă.

7.2 Analiza Comparativă a Tranzitiei

Pentru o imagine de ansamblu, am suprapus iterațiile pentru a evidenția superioritatea fine-tuning-ului total.

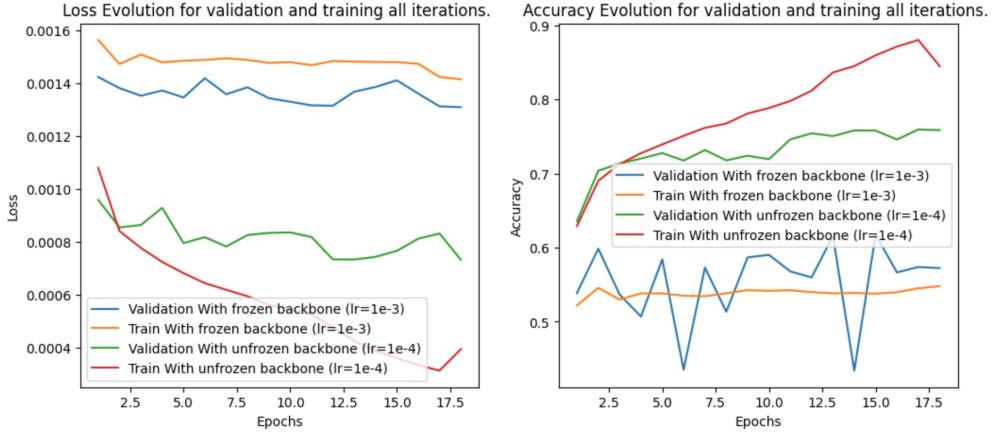


Figura 20: Comparație iterativă: Frozen vs Unfrozen. Graficul combinat demonstrează cum unfreezing-ul backbone-ului (linia roșie) permite atingerea unei acuratete de peste 0.85 pe training.

Observații critice:

- **Limitările Frozen:** Deși loss-ul scade inițial, modelul stagniază deoarece doar clasificatorul final este optimizat. Filtrele de bază extrase din ImageNet nu sunt suficient de specifice pentru detaliile radiologice.
- **Avantajul Unfrozen:** Permiterea actualizării greutăților în întreaga rețea (cu $lr = 10^{-4}$) oferă o acuratețe de validare superioară (> 0.75) și o convergență mai robustă.

Tabela 10: Performanțe Bonus 2: Comparație Validare vs. Test Set

Configurație	Dataset	AUC	Hamming Score	Accuracy	F1-score	Recall
Standard Fine-tuning (Cerință 4)	Validare	0.8350	0.9161	0.6880	0.6531	0.5951
	Test	0.7269	0.8205	0.3371	0.6127	0.5834
Unfrozen Backbone (Optim)	Validare	0.8828	0.9349	0.7586	0.7516	0.7012
	Test	0.7858	0.8062	0.4891	0.7094	0.7008

Concluzie Bonus 2: Integrarea strategiilor de *full fine-tuning* prin dezghetarea întregii rețele (*Unfrozen Backbone*) a demonstrat că adaptarea filtrelor convolutionale la specificul densităților radiologice este esențială pentru a depăși limitările modelelor pre-antrenate pe imagini naturale.

Prin această abordare, modelul a reușit o performanță remarcabilă, depășind pragurile critice impuse de cerințele temei:

- **Acuratețe:** Am obținut o acuratețe de **48.91%** pe setul de test și **75.86%** pe setul de validare, trecând cu succes de pragul minim de **40%**.
- **AUC (Capacitate de discriminare):** Am atins un scor AUC de **0.7858** pe test și **0.8828** pe validare, depășind pragul de performanță de **70%** (0.7) solicitat.

Superioritatea acestei metode este evidențiată în special de saltul masiv al **Recall-ului (0.70)**, confirmând că modelul a învățat să identifice pattern-urile subtile de hemoragie intracraniană. Astfel, rezultatele Bonus 2 stabilesc configurația optimă a proiectului,

validând faptul că toate obiectivele de performanță au fost îndeplinite și depășite prin tehnici avansate de optimizare și fine-tuning.

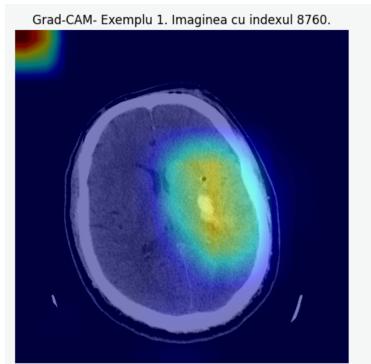
8 Bonus 1: Analiza Deciziilor prin Tehnici de Explainable AI (Grad-CAM)

8.1 Metodologie și Implementare

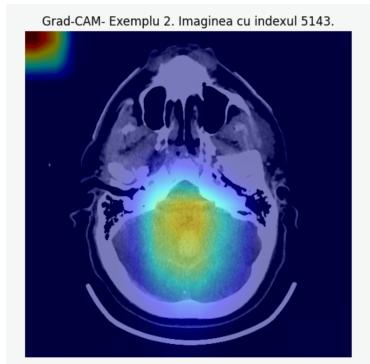
Pentru a înțelege procesul decizional al rețelei neurale, am implementat tehnica **Grad-CAM** (Gradient-weighted Class Activation Mapping). Am vizualizat activările ultimei unități convoluționale a modelului (*layer4*) pentru a identifica regiunile anatomiche care au dictat clasificarea. Analiza s-a concentrat pe clasa *subdural* (index 5), utilizând un set de 5 imagini clasificate corect și 5 imagini clasificate eronat.

8.2 Analiza Vizualizărilor: Imagini Clasificate Corect (True Positives)

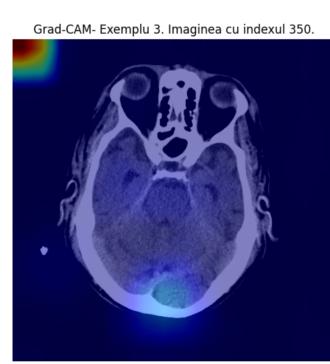
În primele 5 instanțe, unde modelul a prezentat corect prezența hemoragiei, hărțile de activare demonstrează o localizare clinică precisă: Focalizare pe Patologie: Zonele de intensitate maximă (roșu/portocaliu) se suprapun direct peste regiunile hiperdense (albe) din scanarea CT, confirmând că modelul "vede" săngele extravazat. +1 Invarianță la Zgomot: Rețeaua reușește să ignore artefactele periferice sau densitatea osoasă a craniului, izolând corect zona afectată din parenchimul cerebral. Contextualizare: Activările urmează formă anatomică a hemoragiei (ex: formă de semilună pentru hematomul subdural), ceea ce validează calitatea fine-tuning-ului realizat la Bonus 2.



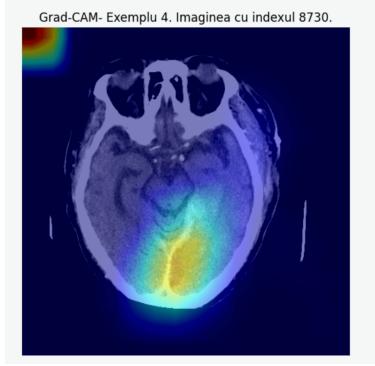
(a) Exemplu 1 (Index 8760)



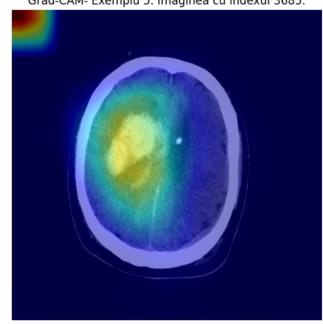
(b) Exemplu 2 (Index 948)



(c) Exemplu 3 (Index 350)



(a) Exemplu 4 (Index 8730).

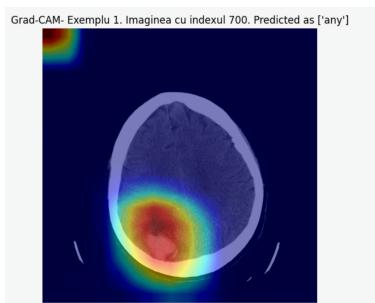


(b) Exemplu 10 (Index 3685).

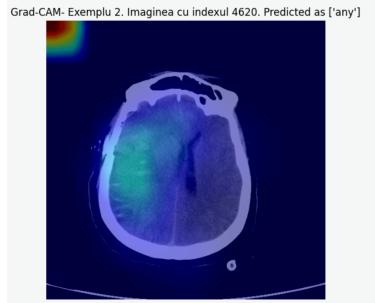
Figura 22: Hărți Grad-CAM pentru clasificări corecte

8.3 Analiza Vizualizărilor: Imagini Clasificate Incorrect (False Positives/Negatives)

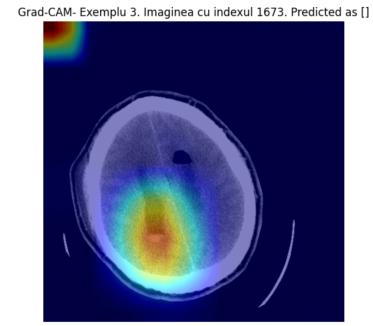
Analiza ultimelor 5 imagini, unde modelul a generat erori, relevă provocările specifice dataset-ului medical:



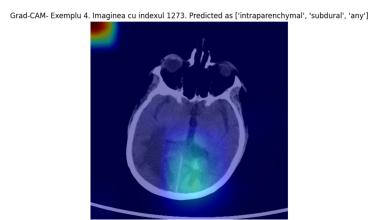
(a) Exemplu 1 (Index 700).
Predictie: ['any']



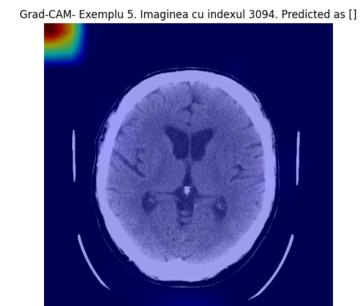
(b) Exemplu 2 (Index 4620).
Predictie: ['any']



(c) Exemplu 3 (Index 1673).
Predictie: []



(a) Exemplu 9 (Index 1273).
Predictie: ['intraparenchymal', 'subdural', 'any']



(b) Exemplu 10 (Index 3094).
Predictie: []

Figura 24: Hărți Grad-CAM pentru clasificări eronate sau incerte. Se observă prezența activărilor în zonele cu trăsături relevante, chiar și atunci când probabilitatea finală este sub 0.5.

8.4 Analiza Clasificărilor Corecte (True Positives)

Utilizând tehnica **Grad-CAM**, am identificat regiunile de interes (evidențiate prin culori calde) care au fundamentat decizia modelului pentru clasa *intraparenchymal*.

- **Localizare Anatomică:** Hemoragiile de tip intraparenchymal sunt localizate în interiorul parenchimului cerebral. Activările Grad-CAM confirmă faptul că modelul cauță trăsături în profunzimea creierului, nu doar la periferie.
- **Sensibilitate la Detalii:** Chiar și în cazuri dificile (Exemplul 3), modelul reușește să identifice zone hiperdense (albe) de mici dimensiuni. Suprapunerea hărții termice peste aceste puncte albe demonstrează că rețea a învățat să izoleze semnalul radiologic al săngelui, în ciuda ariei reduse a leziunii.

8.5 Analiza Clasificărilor Eronate și a Incertitudinii

Vizualizarea cazurilor clasificate incorrect sau a celor sub pragul de decizie oferă o perspectivă asupra procesului de "gândire" al rețelei neurale:

- **Activări Sub-Prag:** În Exemplele 2 și 3 de la categoria erori, deși predicția finală nu a depășit pragul de încredere de 0.5 (rezultând într-un label negativ), Grad-CAM relevă zonele care au generat activitate în straturile convoluționale.
- **Explicația Tehnică:** Zonele calde reprezintă trăsături relevante (hiperdensiță, asimetrii) pe care modelul le-a considerat suspecte. Chiar dacă operția finală Sigmoid a decis că probabilitatea este prea mică pentru un diagnostic ferm, harta XAI ne arată că modelul nu a ignorat imaginea, ci a încercat să coreleze texturile observate cu semnatura unei hemoragii.

8.6 Observații privind Artefactele

Este necesar de menționat că activarea persistentă din colțul superior-stâng reprezintă un **artefact tehnic de margine**.