

Comparació d'Arquitectures de CNN a la Detecció de Malalties Pulmonars usant Radiografies Toràciques

Guillem Simó Hospedales

Grau d'Enginyeria Informàtica – Computació
TFG – Intel·ligència Artificial

Nom Consultor/a

Jordi Delgado Pin

Nom Professor/a responsable de l'assignatura

Susana Acedo Nadal

Data Lliurament

06/2024

Copyright © 2024 GUILLEM SIMÓ
HOSPEDALES.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

FITXA DEL TREBALL FINAL

Títol del treball:	<i>Comparació d'Arquitectures de CNN a la Detecció de Malalties Pulmonars usant Radiografies Toràciques</i>
Nom de l'autor:	<i>Guillem Simó Hospedales</i>
Nom del consultor/a:	<i>Jordi Delgado Pin</i>
Nom del PRA:	<i>Susana Acedo Nadal</i>
Data de lliurament (mm/aaaa):	06/2024
Titulació:	<i>Grau d'Enginyeria Informàtica - Menció en Computació</i>
Àrea del Treball Final:	<i>Intel·ligència Artificial</i>
Idioma del treball:	<i>Català</i>
Paraules clau:	<i>Radiografia, Intel·ligència Artificial, Diagnòstic</i>
Resum del Treball (màxim 250 paraules): Amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball.	
<p>El present treball té com a objectiu comparar diverses arquitectures de xarxes neuronals convolucionals (CNN) en la detecció de malalties pulmonars utilitzant radiografies toràciques. La motivació principal és millorar la precisió i l'eficiència del diagnòstic radiològic, reduint els errors humans i augmentant la velocitat de detecció d'anomalies pulmonars.</p> <p>S'ha realitzat un estudi comparatiu de tres arquitectures de CNN—EfficientNetB0, VGG19 i InceptionV3—aplicant tècniques d'aprenentatge per transferència per adaptar models pre-entrenats al nostre conjunt de dades específic del NIH. La metodologia inclou el pre-processament de dades, l'entrenament dels models amb augmentació de dades, i l'avaluació del rendiment mitjançant diverses mètriques com la precisió, el recall, el F1 Score, l'AUC-ROC i la corba de precisió-recall.</p> <p>L'estudi ha trobat que EfficientNetB0 i InceptionV3 proporcionen un rendiment superior en termes de precisió i capacitat de generalització. Aquestes arquitectures han demostrat un potencial significatiu per millorar el diagnòstic automatitzat de malalties pulmonars. En contrast, mentre que VGG19 ha mostrat un alt recall, presenta una precisió inferior, destacant els reptes en la selecció i optimització de models.</p> <p>Aquesta investigació subratlla el potencial de les CNNs en la imatge mèdica i la importància de seleccionar l'arquitectura de model òptima. Ha estat un repte interessant que ha aportat coneixements valuosos per a futures investigacions en el camp. Millorar i refinar els conjunts de dades i models utilitzats per a l'entrenament serà crucial per a l'avanç continu en el diagnòstic mèdic automatitzat.</p>	

Abstract (in English, 250 words or less):

The present work aims to compare various architectures of convolutional neural networks (CNNs) in the detection of pulmonary diseases using chest X-rays. The primary motivation is to improve the precision and efficiency of radiological diagnosis, reducing human errors and increasing the speed of detecting pulmonary anomalies.

A comparative study of three CNN architectures—EfficientNetB0, VGG19, and InceptionV3—was conducted, applying transfer learning techniques to adapt pre-trained models to our specific dataset from the NIH. The methodology included data pre-processing, training models with data augmentation, and performance evaluation using various metrics such as precision, recall, F1 Score, AUC-ROC, and the precision-recall curve.

The study found that EfficientNetB0 and InceptionV3 provided superior performance in terms of precision and generalization ability. These architectures demonstrated a significant potential to enhance automated pulmonary disease diagnosis. In contrast, while VGG19 showed high recall, it presented lower precision, highlighting the challenges in model selection and optimization.

This research underscores the potential of CNNs in medical imaging and the importance of selecting the optimal model architecture. It was an engaging challenge that contributed valuable insights for future investigations in the field. Improving and refining the datasets and models used for training will be crucial for ongoing advancements in automated medical diagnosis.

Índex

Contenido

1.	Introducció	7
1.1.	Context i justificació del Treball	7
1.2.	Objectius del Treball	7
1.3.	Enfocament i mètode seguit	8
1.4.	Planificació del Treball	9
1.5.	Breu descripció dels altres capítols de la memòria.....	11
2.	Estat de l'Art	12
2.1.	Aprenentatge automàtic (<i>Machine Learning</i>)	12
2.2.	Algoritmes de classificació	13
2.3.	Aprenentatge profund (<i>Deep Learning</i>)	15
2.4.	Xarxes neuronals convolucionals (<i>Convolutional Neural Networks</i>)	17
2.5.	Aprenentatge per transferència (<i>Transfer Learning</i>).....	18
2.6.	Ús de Models Pre-entrenats a Xarxes Neuronals Convolucionals (CNN)	19
2.7.	InceptionV3.....	20
2.8.	VGG19	21
2.9.	EfficientNet-b0.....	22
2.10.	CNNs al sector sanitari	23
2.11.	CNN a l'anàlisi d'imatges de raigs X.....	24
3.	Descripció detallada del treball pràctic.....	26
3.1.	Entorn utilitzat	26
3.2.	Base de dades emprada	26
3.3.	Pre-processament de dades.....	27
3.4.	Entrenament dels diferents models	28
4.	Descripció y anàlisi dels resultats	33
5.	Conclusions	39
6.	Bibliografia	40

Llista de figures

Figura 1: Planificació, Part1	10
Figura 2: Planificació, Part 2	10
Figura 3: Classificació gràfica del Machine Learning [7].....	12
Figura 4: procés d'entrenament-validació-test [10]	14
Figura 5: Neurones Biològiques y Artificials [17].....	15
Figura 6: Representació de les capes d'una xarxa neuronal [18].....	15
Figura 7: Xarxa Neuronal durant el procés d'aprenentatge [19]	16
Figura 8: Esquema d'una CNN [27]	17
Figura 9: Representació gràfica del Transfer Learning [37].....	19
Figura 10: Arquitectura InceptionV3 [42].....	20
Figura 11: Representació gràfica de VGG19 [48]	21
Figura 12: Exemple gràfic d'EfficientNetB0 [51].....	22
Figura 13: Estructura de les dades d'entrenament	27
Figura 14: Verificació de la distribució d'imatges d'entrenament	28
Figura 15: Generador de dades d'imatge.....	29
Figura 16: Càrrega de dades al generador d'entrenament	30
Figura 17: Definició i compilació d'EfficientNetB0	31
Figura 18: Codi del Model Checkpoint	31
Figura 19: Codi del Early Stopping.....	32
Figura 20: Codi de reducció del Learning Rate.....	32
Figura 21: Funció d'avaluació dels models	33
Figura 22: Comparació de la precisió i pèrdua de validació per EfficientNetB0, VGG19 i InceptionV3.....	34
Figura 23: Matriu de confusió de EfficientNetB0, VGG19 i InceptionV3.....	35
Figura 24: Corba ROC per EfficientNetB0, VGG19 i InceptionV3	36
Figura 25: Corba de Precisió-Recall per EfficientNetB0, VGG19 i InceptionV3	37

1. Introducció

1.1. Context i justificació del Treball

Punt de partida del treball:

El punt de partida d'aquest treball és la necessitat de millorar el procés de diagnòstic mèdic a través de l'anàlisi de radiografies toràciques. Actualment, el diagnòstic radiològic pot ser subjectiu i estar subjecte a errors humans, cosa que pot portar a diagnòstics inexactes i retards en el tractament adequat dels pacients. Tot i els avenços en tecnologia mèdica, la identificació precisa d'anomalies en radiografies, com patologies pulmonars, continua sent un desafiament per als professionals de la salut.

Justificació de la rellevància del tema:

El tema és rellevant tant per a la societat com per a l'àmbit científic degut al seu impacte en la salut i el benestar de les persones. La detecció primerenca i precisa de malalties pulmonars a través de radiografies és crucial per a un tractament efectiu i millors resultats per als pacients. Millorar la precisió i l'eficiència del diagnòstic radiològic no només pot salvar vides, sinó també reduir costos d'atenció mèdica i millorar la qualitat de vida dels pacients.

Aportació realitzada:

L'objectiu principal d'aquest treball és desenvolupar un sistema d'Intel·ligència Artificial (IA) que sigui capaç de detectar i classificar anomalies pulmonars en radiografies toràciques amb alta precisió. L'aportació consisteix a comparar l'eficiència de diferents models de xarxes neuronals (*EfficientNet*, *VGG19* i *InceptionV3*), mitjançant el *transfer learning*, per trobar-ne el més òptim. No es pretén trobar una solució definitiva al problema, sinó facilitar el camí per a investigacions futures en seleccionar les eines més adequades per dur a terme aquest tipus d'anàlisi.

1.2. Objectius del Treball

- Investigar i revisar la literatura científica relacionada amb l'ús d'intel·ligència artificial en l'anàlisi de radiografies toràciques per al diagnòstic mèdic.
- Desenvolupar un sistema de detecció amb diferents models de CNN (*EfficientNet*, *VGG19* i *InceptionV3*) per veure quin s'adapta millor al nostre cas.
- Avaluar i comparar l'eficàcia i la precisió dels models mitjançant una sèrie de mètriques.
- Elaborar una guia clara i concisa per a futures investigacions sobre aquest camp.

1.3. Enfocament i mètode seguit

L'enfocament d'aquest treball consisteix a fer servir tècniques de transferència de coneixement, aprenentatge automàtic i xarxes neuronals per trobar l'enfocament més eficient a l'hora d'analitzar imatges de radiografies toràciques. Així mateix, es realitzarà un procés de recerca sobre altres estudis semblants i es documentaran tots els passos.

Estratègia:

- **Revisió de literatura científica:** revisió exhaustiva de la literatura científica relacionada amb l'ús d'IA en l'anàlisi de radiografies per al diagnòstic mèdic. Això servirà com a base per comprendre l'estat de l'art, identificar enfocaments i tècniques prometedores i establir fonaments teòrics sòlids per al desenvolupament del sistema d'IA.
- **Desenvolupament del Sistema d'IA:** Mitjançant el mètode de transferència de coneixement, s'implementaran i s'adaptaran tres models de CNN pre-entrenats (*EfficientNet*, *VGG19* i *InceptionV3*) al nostre cas d'ús.
- **Avaluació de l'Eficàcia i Precisió:** Un cop desenvolupats els models, es procedirà a avaluar-ne l'eficiència i precisió per separat i s'avaluarà quin és més adequat per a aquest cas basant-nos en mètriques com el temps de desenvolupament, entrenament, velocitat de processament i precisió dels resultats.
- **Documentació del procés:** Es desenvoluparà una guia de tot el procés per facilitar l'accés a aquelles persones interessades en aquest camp, fomentant futures investigacions i aplicacions pràctiques al sector de la salut.

Justificació de l'Estratègia:

Aquesta estratègia s'ha considerat la més apropiada per assolir els objectius per les raons següents:

- **Ampli Fonament Teòric:** La revisió exhaustiva de la literatura científica permetrà establir una base teòrica sòlida per al desenvolupament del sistema d'IA, aprofitant els avenços i enfocaments reeixits prèviament investigats.
- **Desenvolupament Pràctic del Sistema d'IA:** L'adaptació de models pre-entrenats permetrà reduir el temps de desenvolupament així com adaptar i personalitzar les tècniques i algoritmes segons les necessitats específiques del diagnòstic radiològic, maximitzant l'eficàcia i la precisió del sistema.
- **Avaluació Rigorosa:** Realitzar una anàlisi completa per determinar quin enfocament és més eficient per a aquest cas d'ús és fonamental per assolir els objectius del treball.

1.4. Planificació del Treball

El pla de treball s'estructura en cinc PACs (Proves d'Avaluació Contínua), cadascuna que representa una etapa clau en la realització del treball final de grau (TFG).

- Primera PAC (12 - 26 de març): Elaboració i entrega del Pla de Treball. Definició del tema i objectius, investigació de l'estat de l'art i contextualització del camp d'estudi, documentació de tecnologies relacionades i selecció de la base de dades que cal utilitzar.
- Segona PAC (27 de març – 3 de maig): Desenvolupament del treball, part 1. Lliurament de la tecnologia i la base de dades proposades per al TFG, juntament amb l'inici del desenvolupament de la part pràctica.
- Tercera PAC (4 - 29 de maig): Desenvolupament del treball, part 2. Anàlisi del feedback rebut a la PEC2, avaluació de possibles canvis en l'enfocament del treball, i finalització de la part pràctica.
- Quarta PAC (30 de maig – 16 de juny): Redacció de la memòria del TFG. Documentació del treball realitzat, incloent-hi la introducció, metodologia, resultats i conclusions.
- Cinquena PAC (17-30 de juny): Preparació de la presentació i defensa del TFG. Preparació de la presentació per a la defensa davant d'un tribunal avaluador.

A la pàgina següent es pot observar el diagrama de Gantt corresponent a aquesta planificació representat en dues il·lustracions.

	Marzo															Abril																																						
	PEC1															PEC2 (Parte 1 del trabajo)																																						
	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30				
Elaboración y entrega del Plan de Trabajo (PEC1)																																																						
Estado del Arte y contextualización del campo de estudio																																																						
Documentación e Investigación de tecnologías relacionadas																																																						
Selección de la Base de Datos																																																						
Entrega de la tecnología propuesta para la realización del TFG																																																						
Entrega de Base de Datos propuesta para la realización del TFG																																																						
Desarrollo de la parte práctica del TFG																																																						

Figura 1: Planificació, Part 1

	Mayo															Junio																																													
	PEC2			PEC3 (Parte 2 del trabajo)															PEC4 (MEMORIA)															PEC5 (Presentación)																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Desarrollo de la parte práctica del TFG																																																													
Entrega de la PEC2 (Desarrollo del Trabajo, Parte 1)																																																													
Calificación PEC2																																																													
Análisis del feedback de la PEC2																																																													
Valorar si es necesario cambiar el enfoque y aplicar cambios																																																													
Entrega de la PEC3 (Desarrollo del Trabajo, Parte 2)																																																													
Calificación PEC3																																																													
Análisis del feedback de la PEC3																																																													
Redacción de la Memoria																																																													
Entrega de la PEC4 (Memoria)																																																													
Preparación de la Presentación y Defensa																																																													
Entrega de la PEC5a (Preparación)																																																													
PEC5b (Defensa) [Fecha por determinar]																																																													

Figura 2: Planificació, Part 2

1.5. Breu descripció dels altres capítols de la memòria

Capítol 2: Estat de l'Art

Aquest capítol presenta una revisió exhaustiva de la literatura científica sobre l'aprenentatge automàtic, l'aprenentatge profund i l'ús de xarxes neuronals convolucional (CNN) en la detecció de malalties pulmonars. Proporciona el context teòric necessari per comprendre els fonaments del treball.

Capítol 3: Descripció detallada del treball pràctic

Aquí es detallen els aspectes pràctics de la implementació del treball, incloent-hi l'entorn utilitzat, la base de dades emprada, el pre-processament de dades, i l'entrenament dels diferents models CNN. Aquest capítol explica els passos seguits per desenvolupar el sistema d'anàlisi de radiografies toràciques.

Capítol 4: Descripció i anàlisi dels resultats

Es presenten els resultats obtinguts de l'entrenament i l'avaluació dels models CNN. Es discuteixen les mètriques de rendiment, com la precisió, el recall, el F1 Score, l'AUC-ROC, i la corba de precisió-recall, i es comparen les diferents arquitectures per determinar la més òptima.

Capítol 5: Conclusions

Aquest capítol ofereix una avaluació final del projecte, destacant les conclusions principals obtingudes, comparant les expectatives inicials amb els resultats finals. També es discuteixen les possibles línies de treball futur i les implicacions del treball realitzat.

Capítol 6: Bibliografia

Proporciona una llista de les fonts i referències utilitzades per a la investigació i el desenvolupament del projecte. Aporta el suport teòric necessari i mostra l'amplitud de la recerca realitzada.

2. Estat de l'Art

2.1. Aprentatge automàtic (*Machine Learning*)

Al llarg de la segona meitat del segle XX, moltes empreses, associacions i investigadors es van adonar que cada vegada comptaven amb una major quantitat de dades relacionades amb els seus respectius camps, fent impossible per a un humà processar tota aquesta informació. Va ser llavors quan es va començar a gestar la idea de crear un programa que fos capaç de processar i analitzar totes aquestes dades.

A grans trets, i com bé indica el seu nom, l'aprenentatge automàtic (en endavant *machine learning*, pel seu nom en anglès) consisteix en una sèrie de processos, operacions i algorismes que permeten que un programa aprengui a prendre decisions sobre la base d'una entrada de dades de forma automàtica. Per això, en primer lloc, cal comptar amb un gran volum de dades. Seguidament, es prenen aquestes dades i s'introdueixen al programa, on són processats per un algorisme i estigui presa una decisió. Finalment, aquesta decisió és revisada per comprovar-ne la veracitat. Aquest procés es repeteix una vegada i una altra durant diverses iteracions fins a obtenir el percentatge d'encert desitjat.

Un dels problemes més grans en desenvolupar un sistema d'aprenentatge automàtic és l'aparició de biaixos, una tendència errònia en analitzar les dades. Això sol passar quan la base de dades no està prou balancejada. Tenir moltes dades no serveix de res si no representen equitativament tots els casos que volem analitzar. Un conjunt de dades mal balancejat pot resultar en un índex de predicció alt, però considerablement inferior per a casos amb dades insuficients.

Un cas molt conegut sobre el biaix és el dels problemes que tenen els programes de detecció facial per detectar persones d'ètniques no caucàsiques. Això és degut, en gran mesura, al fet que aquests han estat entrenats amb imatges de persones amb la pell clara.

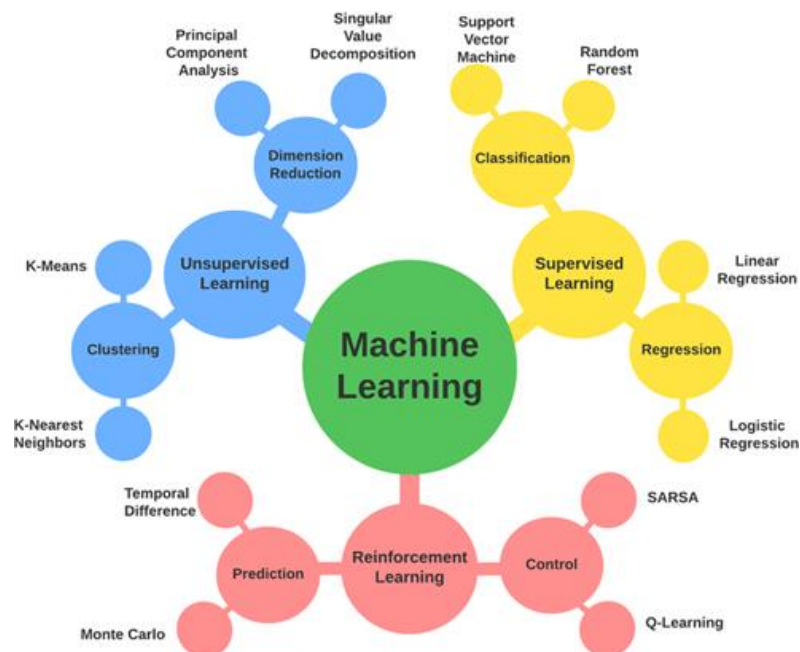


Figura 3: Classificació gràfica del Machine Learning [7]

Veiem que, en funció del seu comportament i mode de processament, podem classificar aquests models en dos tipus:

- **Aprenentatge supervisat:** aquests models usen conjunts de dades etiquetades, fet que facilita que el mateix model avalui la seva eficàcia i millori amb el temps, ja que pot comprovar automàticament si els resultats obtinguts es corresponen a l'etiquetatge esperat. L'aprenentatge supervisat es pot classificar en dos tipus:
 - **Classificació:** s'utilitza per predir categories discretes, aprenent a assignar etiquetes de classes a les entrades.
 - **Regressió:** es fa servir per predir valors continus, aprenent a estimar un valor basat en les entrades.
- **Aprenentatge no supervisat:**
 - **Agrupament o *clustering*:** agrupa les dades sobre la base de la similitud.
 - **Associació:** es fa servir per descobrir regles en grans conjunts de dades, trobant relacions entre variables.

2.2. Algoritmes de classificació

Els algorismes de classificació són una forma d'aprenentatge supervisat on el resultat o la predicció pertanyen a una categoria específica. Aquests algorismes s'utilitzen per determinar la classe a què pertany una entrada. Hi ha algorismes de classificació binària, que prediuen entre dues possibles classes, com identificar si un correu és o no *spam*. També hi ha algorismes de classificació amb múltiples classes que prediuen entre més de dues categories, com classificar diferents tipus de fruites segons la seva espècie.

Igual que tots els mètodes d'aprenentatge supervisat, els algorismes de classificació requereixen dades etiquetades per al seu entrenament. Aquestes dades han de ser representatives del problema que es pretén resoldre, de manera que el model, després del seu entrenament, pugui fer prediccions correctes sobre dades noves que no ha vist abans. Quan el model ho aconsegueix, es diu que té capacitat de generalització. És per això que l'etiquetatge de les dades s'ha de fer meticulosament, ja que un error en aquest procés podria provocar que el model no funcioni correctament.

El desenvolupament d'algorismes de classificació a l'aprenentatge supervisat involucra diverses fases crucials, cadascuna amb un propòsit específic. Aquestes fases inclouen entrenament, validació i test del model. A continuació, s'expliquen detalladament aquestes fases, així com la proporció típica de dades utilitzada en cadascuna i la seva importància.

Fase d'Entrenament

A la fase d'entrenament, el model aprèn a partir d'un conjunt de dades etiquetades. Aquestes dades contenen exemples d'entrada i les sortides o etiquetes corresponents. Durant aquesta fase, l'algorisme ajusta els paràmetres interns per minimitzar l'error en les prediccions fetes sobre les dades d'entrenament. Generalment, es fa servir entre el 60% i el 80% del total de dades disponibles per a l'entrenament.

Aquesta fase és crucial perquè és on el model "aprèn" les característiques i patrons de les dades. Si el conjunt de dades d'entrenament és representatiu del problema real, el model pot generalitzar bé i fer prediccions precises sobre noves dades. Sense un bon entrenament, el model no serà capaç d'aprendre adequadament i el rendiment serà pobre.

Fase de Validació

La fase de validació utilitza un conjunt de dades diferent al d'entrenament per avaluar el rendiment del model mentre s'ajusta durant la fase d'entrenament. Aquest conjunt de dades es fa servir per afinar els híper-paràmetres del model, com ara la taxa d'aprenentatge, el nombre de capes en una xarxa neuronal, etc.

Permet identificar problemes com el sobre-ajustament i sub-ajustament, proporcionant una indicació més realista del rendiment del model en dades no vistes. Els ajustaments realitzats basant-se en les dades de validació poden millorar significativament la capacitat del model per generalitzar. Sol emprar-se entre el 10% i el 20% del total de dades disponibles per a la validació.

Fase de Test

A la fase de test, s'utilitza un altre conjunt de dades, que no s'han utilitzat ni per a l'entrenament ni per a la validació, per avaluar el rendiment final del model. Aquest conjunt de dades serveix com un indicador objectiu de com el model probablement s'exercirà al món real. Generalment, es reserva entre el 10 i el 20% del total de dades disponibles per al test.

La fase de test és crítica per obtenir una estimació objectiva i final del rendiment del model. Assegura que el model no només funcioni bé amb les dades en què va ser entrenat i validat, sinó que també pot generalitzar dades completament noves, ajudant a entendre si el model és viable per a la seva implementació en un entorn real.

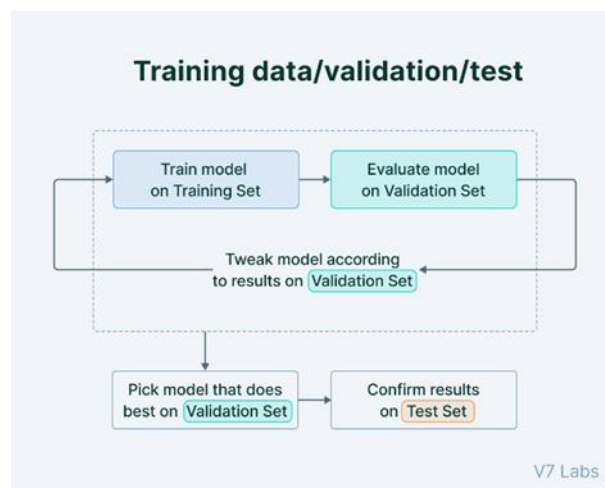


Figura 4: procés d'entrenament-validació-test [10]

2.3. Aprentatge profund (*Deep Learning*)

Introducció

L'aprenentatge profund (en endavant *Deep Learning*, pel seu nom en anglès) és una subdisciplina del *Machine Learning* que utilitza xarxes neuronals profundes per modelar i resoldre problemes complexos. Aquestes xarxes es basen en l'estructura i el funcionament del cervell humà, permetent aprendre representacions jeràrquiques de les dades a través de múltiples capes de processament.

Funcionament de les Neurons al *Deep Learning*

Cada neurona a les capes de la xarxa neuronal realitza un processament específic, rebent múltiples entrades, cadascuna ponderada per un coeficient que en reflecteix la importància. Aquestes entrades ponderades es sumen i el resultat passa per una funció d'activació no lineal, com ara la funció sigmoide, la tangent hiperbòlica o la ReLU (*Rectified Linear Unit*). La funció d'activació introdueix no-linealitats en el model, permetent a la xarxa aprendre i representar relacions complexes en les dades.

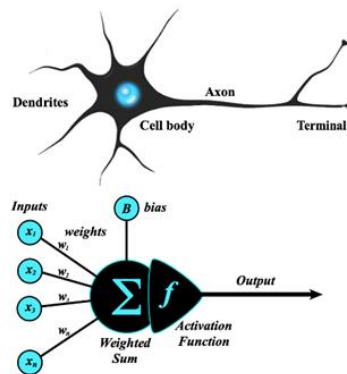


Figura 5: Neurons Biològiques y Artificials [17]

Arquitectura d'una xarxa neuronal

Les xarxes neuronals consisteixen en diverses capes de neurones artificials, com les descrites a l'apartat anterior. Aquestes capes inclouen:

- Capa d'entrada: rep les dades d'entrada, ja siguin textos, imatges o qualsevol altre tipus.
- Capes Ocultes: processen les entrades aplicant transformacions no lineals. Permeten que la xarxa aprengui característiques complexes de les dades.
- Capa de sortida: produeix la predicció final.

Cada neurona en aquestes capes pondera les entrades, aplica una funció d'activació i produeix una sortida.

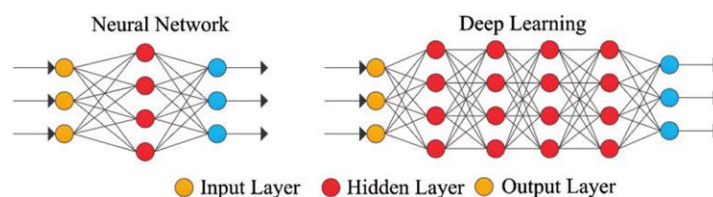


Figura 6: Representació de les capes d'una xarxa neuronal [18]

Procés d'entrenament

L'entrenament d'una xarxa neuronal profunda implica:

- *Forward Propagation*: les entrades es propaguen cap endavant a través de les capes de la xarxa, aplicant funcions d'activació en cada capa, fins arribar a la capa de sortida, produint una predicció basada en els pesos actuals de les connexions.
- *Backward Propagation*: l'error de la predicció es calcula comparant-lo amb la sortida desitjada utilitzant una funció de pèrdua i aquest es propaga enrere a través de les capes, ajustant els pesos de les connexions per reduir l'error en prediccions futures. Aquest ajustament es realitza mitjançant el mètode de descens del gradient, que actualitza els pesos en la direcció oposada al gradient de la funció de pèrdua respecte als pesos, amb una magnitud determinada per la taxa d'aprenentatge.

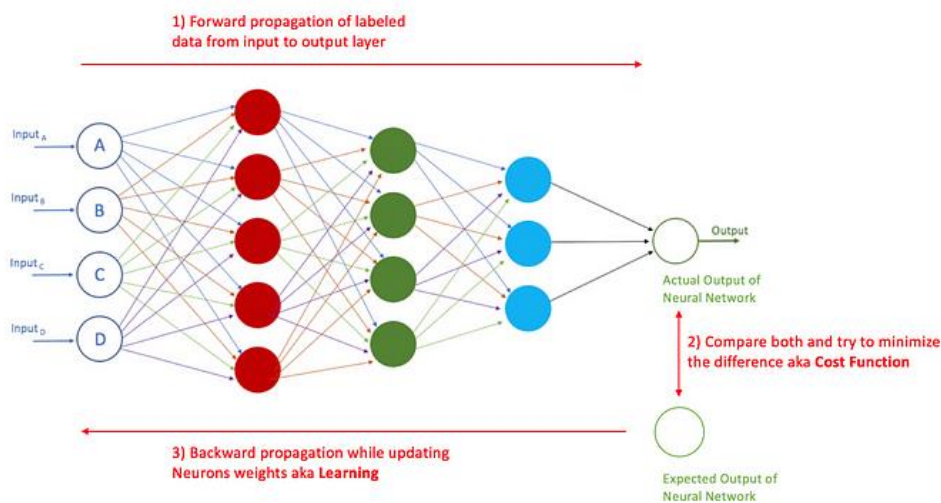


Figura 7: Xarxa Neuronal durant el procés d'aprenentatge [19]

Aplicacions i Avantatges

Les xarxes profundes poden aprendre representacions complexes sense necessitat d'intervenció manual, cosa que les fa especialment útils a:

- Reconeixement d'imatges: classificació d'imatges i detecció d'objectes. Per exemple, el model AlexNet va establir un nou estàndard en la classificació d'imatges al guanyar la competència ImageNet el 2012 (Krizhevsky, Sutskever, & Hinton, 2012) [15].
- Processament del Llenguatge Natural: Traducció automàtica i anàlisi de sentiments. Un exemple destacat és el sistema de traducció de Google Translate, que utilitza xarxes neuronals profundes per traduir textos entre diversos idiomes amb alta precisió (Wu et al., 2016) [20].

2.4. Xarxes neuronals convolucional (Convolutional Neural Networks)

Introducció

Les Xarxes Neuronals Convolucional (en endavant *CNN*, per les seves sigles en anglès) són un tipus específic de xarxa neuronal dissenyada per processar dades amb una estructura de quadrícula, com ara imatges i vídeos. Les CNN són altament efectives en la detecció i el reconeixement de patrons espacials i temporals.

Estructura de les CNNs

Les CNNs es componen de diversos tipus de capes especialitzades:

- Capes Convolucional: apliquen filtres per detectar característiques locals, com ara vores i textures. Aquestes capes conserven la relació espacial entre els píxels de la imatge.
- Capes de *Pooling*: redueixen les dimensions de les dades resumides, permetent la detecció de característiques en escales més petites.
- Capes Completament Connectades: Aquestes capes, situades al final de la xarxa, realitzen la classificació final basada en les característiques extrems.

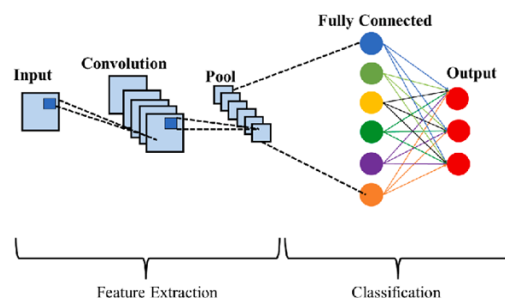


Figura 8: Esquema d'una CNN [27]

Procés d'Entrenament

L'entrenament d'una CNN segueix els mateixos principis bàsics que altres xarxes neuronals:

- Forward Propagation: Les imatges d'entrada es processen a través de les capes convolucional i de *pooling*, extraient-ne característiques.
- Backward Propagation: L'error de la predicció es retro-propaga per ajustar els pesos de les connexions fent servir el descens del gradient.

Aplicacions i Avantatges

Les CNN són particularment útils a:

- Visió per Ordinador: Classificació d'imatges, detecció d'objectes i reconeixement facial. Per exemple, les CNN s'han utilitzat amb èxit en sistemes per detectar caigudes de persones basant-se en imatges de càmeres de vigilància (Jian et al., 2018) [28].
- Reconeixement de Patrons: Anàlisi de seqüències temporals en vídeos i sèries de temps. Un exemple és l'ús de CNN per a l'anàlisi de moviments en vídeos per detectar activitats humanes (El-Ghaish et al., 2018) [29].

2.5. Aprenentatge per transferència (*Transfer Learning*)

Introducció

Entrenar una xarxa neuronal des de zero requereix una quantitat significativa de dades i temps de computació, sovint estenent-se a dies o setmanes. La tècnica de transferència de coneixement (en endavant *Transfer Learning*, pel seu nom en anglès) permet utilitzar una xarxa prèviament entrenada i incorporar-la al nostre model. Aquest enfocament aprofita l'entrenament ja realitzat en un altre model, cosa que significa que els paràmetres ja estan ajustats, i es continua l'entrenament amb el nostre conjunt de dades específic.

Fonaments del *Transfer Learning*

Una xarxa neuronal aprèn diferents nivells de detall a cadascuna de les seves capes. Les primeres capes aprenen característiques més generals, mentre que les capes posteriors aprenen característiques més específiques. Si les dades originals utilitzades per entrenar la xarxa són àmplies i generals, les característiques apreses poden ser reutilitzades per a molts problemes diferents dins de l'àmbit de la visió per ordinador.

Mètodes del *Transfer Learning*

Hi ha dos mètodes principals per aplicar *Transfer Learning*:

Feature Extraction:

- Aquest mètode utilitza les característiques apreses per la xarxa pre-entrenada per extreure característiques de les noves dades. Després, aquestes característiques es processen mitjançant les darreres capes del classificador, que s'entrenen des de zero.
- Les CNN es componen de dues parts principals: la base convolucional i el classificador completament connectat (*fully-connected*). *Feature Extraction* reutilitza la base convolucional de la xarxa pre-entrenada i entrena des de zero la part *fully-connected* del classificador.
- Aquest mètode és recomanable quan es disposa d'un conjunt de dades petit i semblant al conjunt de dades del model pre-entrenat.

Fine-Tuning:

- Realitza un ajustament més fi que *Feature Extraction*. En comptes de mantenir congelades les capes de la base convolucional, s'entrena algunes de les capes finals juntament amb el classificador *fully-connected*.
- Aquest ajustament permet que les representacions més abstractes del model pre-entrenat s'ajustin millor al nostre conjunt de dades específic.
- És particularment útil quan el conjunt de dades amb què es treballa és diferent del conjunt de dades original amb què es va entrenar la xarxa, permetent un ajust més detallat i específic. No obstant, l'entrenament addicional implica un major temps d'execució.

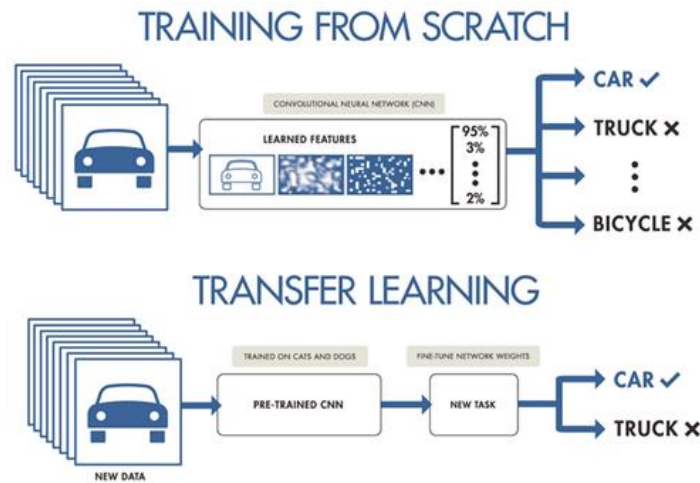


Figura 9: Representació gràfica del Transfer Learning [37]

Avantatges del Transfer Learning

El benefici principal del *Transfer Learning* és que permet entrenar models d'alt rendiment amb conjunts de dades més petits i menys recursos computacionals. Això és especialment útil en aplicacions on recol·lectar grans quantitats de dades etiquetades és costós o inviable.

2.6. Ús de Models Pre-entrenats a Xarxes Neuronals Convolucionals (CNN)

Introducció

Seguint l'explicació del *transfer learning*, l'ús de models pre-entrenats a les CNN ha revolucionat el camp de la visió per ordinador. Aquests models s'entrenen prèviament en grans conjunts de dades, com ara *ImageNet*, que conté milions d'imatges classificades en milers de categories. Posteriorment, es reutilitzen aquests models per a tasques específiques, aprofitant les característiques i els patrons apresos durant el pre-entrenament. Aquest enfocament permet estalviar temps i recursos computacionals significatius i millorar la precisió i l'eficiència del model.

Implementació de Models Pre-entrenats

Els models pre-entrenats funcionen mitjançant un procés de *Transfer Learning*, que inclou dos mètodes principals: *Feature Extraction* i *Fine-Tuning*, els quals ja hem explicat anteriorment. En el següent apartat, profunditzarem com s'implementen aquests mètodes utilitzant arquitectures populars com ara *InceptionV3*, *VGG* i *EfficientNet-B0*.

2.7. InceptionV3

InceptionV3 és un model de CNN desenvolupat per Google, presentat a "*Rethinking the Inception Architecture for Computer Vision*" el 2015 (Szegedy et al., 2016) [43]. Es destaca pel seu ús eficient de recursos computacionals i la precisió en el reconeixement d'imatges.

Arquitectura

- Blocs *Inception*: Permeten capturar informació a múltiples escales mitjançant l'aplicació de filtres de diferents mides (1x1, 3x3, 5x5) i la concatenació dels resultats.
- Factorització de Convolucions: Redueix la quantitat de paràmetres i el cost computacional en descompondre convolucions més grans en convolucions més petites i eficients.
- Convolucions Asimètriques: Descomponen una convolució NxN en una combinació de convolucions 1xN i Nx1 per millorar l'eficiència.
- Reducció i projecció: utilitzen convolucions 1x1 per reduir la dimensionalitat abans d'aplicar convolucions més complexes.

Aquesta arquitectura avançada permet a *InceptionV3* ser eficient i efectiu en la captura de patrons complexos en imatges.

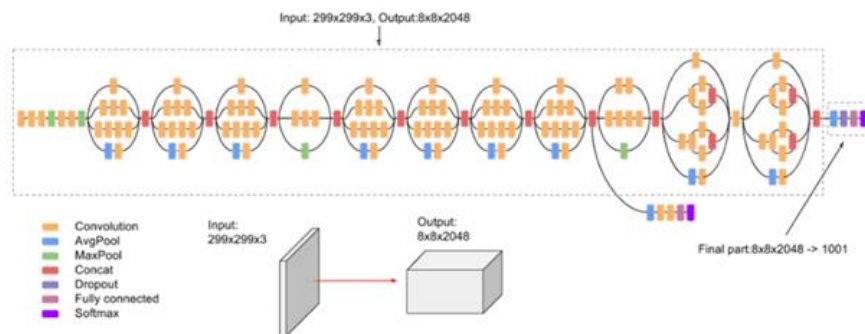


Figura 10: Arquitectura InceptionV3 [42]

Avantatges i Aplicacions

- Eficiència Computacional: Equilibri òptim entre profunditat i eficiència computacional.
- Rendiment superior: alt rendiment en *benchmarks* de visió per ordinador, fent-lo molt eficient en tasques com el reconeixement i classificació d'imatges.
- Aplicacions: Classificació d'imatges, detecció i segmentació d'objectes i aprenentatge per transferència en tasques relacionades.

Limitacions

- Complexitat Arquitectònica: és més complex que models com *VGG19*, cosa que en pot dificultar la implementació.
- Requisits computacionals: tot i la seva eficiència, continua sent demandant en recursos computacionals.

2.8. VGG19

VGG19 és un model de CNN desenvolupat pel Grup de Geometria Visual (VGG, per les seves sigles en anglès) de la Universitat d'Oxford, presentat al "Desafiament de Reconeixement Visual a Gran Escala de *ImageNet*" (Russakovsky et al., 2014) [47]. Aquest model compta amb 19 capes, incloses 16 capes convolucionals i 3 capes completament connectades, i utilitza filtres petits de 3x3 per capturar detalls fins a les imatges.

Arquitectura

- Capes Convolucionals: Utilitza filtres de 3x3 a totes les capes, seguits d'activacions *ReLU*.
- Capes de *Pooling*: Empra capes de *max-pooling* amb una finestra de 2x2 i un pas de 2 per reduir les dimensions espacials i controlar el sobre-ajustament.
- Capes Completament Connectades: Tres capes al final, amb les dues primeres tenint 4096 canals i la darrera realitzant la classificació amb *softmax*.

Aquesta estructura permet a VGG19 aprendre característiques complexes i representacions jeràrquiques de les imatges (Simonyan i Zisserman, 2014) [46].

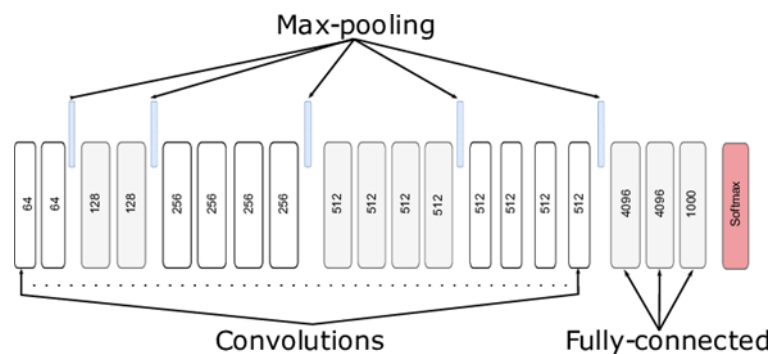


Figura 11: Representació gràfica de VGG19 [48]

Avantatges i Aplicacions

- Profunditat i Uniformitat: La profunditat i l'ús uniforme de filtres de 3x3 faciliten la implementació i la comprensió del model.
- Rendiment: Destacat a ILSVRC 2014, demostrant alta precisió en la classificació d'imatges.
- Aplicacions: Classificació d'imatges, detecció i segmentació d'objectes i aprenentatge per transferència a diversos dominis.

Limitacions

- Intensiu en Computació: El gran nombre de paràmetres dels models VGG és un desavantatge. VGG-16 i VGG-19 són exigents a nivell computacional a causa dels seus respectius 138 milions i 144 milions de paràmetres. (A Deep Dive into Pretrained Models: VGG-16, VGG-19, ResNet, AlexNet, and Inception, 2023) [39].

2.9. EfficientNet-b0

EfficientNetB0 és el model base de la família *EfficientNet*, desenvolupada per Google i presentada a l'article "*EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*" el 2019 [50]. Aquest model es destaca per la capacitat d'aconseguir un alt rendiment en tasques de classificació d'imatges mantenint una eficiència computacional excepcional.

Arquitectura

EfficientNetB0 introdueix un enfocament innovador per escalar CNNs, combinant tres dimensions d'escalat: profunditat, amplitada i resolució.

- Profunditat: Augmenta el nombre de capes a la xarxa, permetent aprendre característiques més complexes.
- Amplitada: Incrementa el nombre de canals a cada capa, cosa que permet capturar més informació detallada.
- Resolució: millora la resolució de les imatges d'entrada, tot preservant més detalls de les característiques espacials.

El bloc bàsic d'*EfficientNetB0* és el MBConv (Mobile Inverted Bottleneck Convolutional Block), que inclou:

- Convolucions de profunditat: apliquen convolucions separades a cada canal d'entrada, reduint el nombre de paràmetres i millorant l'eficiència computacional.
- Convolucions d'Expansió: Augmenten el nombre de canals abans d'aplicar la convolució de profunditat i permeten capturar característiques més complexes.
- Convolucions Invertides: Utilitzen una configuració en què l'expansió passa abans de la convolució, seguida d'una reducció de dimensionalitat.

A més, *EfficientNetB0* utilitza la funció d'activació "swish", que millora la convergència i el rendiment del model (Tan i Le, 2019) [50].

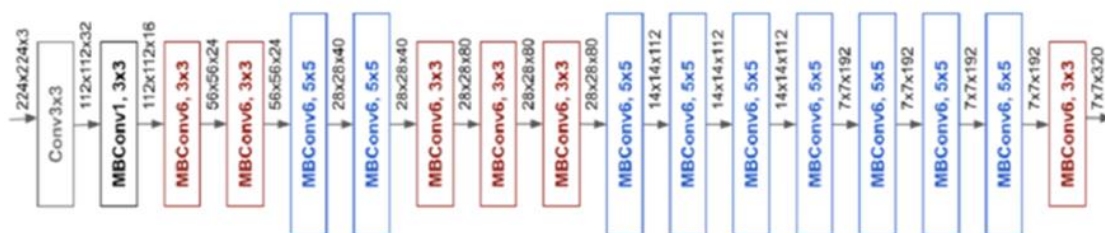


Figura 12: Exemple gràfic d'*EfficientNetB0* [51]

Avantatges i Aplicacions

- **Eficiència Computacional:** *EfficientNetB0* està dissenyat per ser altament eficient en termes de còmput i memòria, aconseguint un equilibri òptim entre precisió i cost computacional.
- **Rendiment Superior:** *EfficientNetB0* ha demostrat un rendiment d'avantguarda en diversos *benchmarks* de visió per ordinador, superant models anteriors amb menys paràmetres i menys *FLOPs* (Operacions de Punt Flotant per Segon).
- **Escalabilitat:** L'enfocament d'escalat compost d'*EfficientNet* permet que la família de models s'ajusti a necessitats diverses de recursos computacionals i precisió, des de dispositius mòbils fins a grans centres de dades.

EfficientNetB0 s'ha utilitzat àmpliament en diverses aplicacions:

- **Classificació d'imatges:** Com a model pre-entrenat amb *ImageNet*, és una opció popular per a tasques de classificació d'imatges en una varietat de dominis.
- **Detecció i Segmentació d'Objectes:** L'eficiència i la precisió d'*EfficientNetB0* la fan adequada per a tasques de detecció i segmentació d'objectes en temps real.
- **Aprentatge per Transferència:** S'utilitza sovint com a punt de partida per entrenar models en tasques relacionades, aprofitant-ne els pesos pre-entrenats per millorar l'eficiència de l'entrenament i el rendiment.

Limitacions

- **Complexitat del Disseny:** Encara que *EfficientNetB0* és eficient, el seu disseny i configuració poden ser més complexos en comparació amb models més simples com *ResNet* o *VGG*, cosa que en pot dificultar la seva implementació i ajust.
- **Requisits d'Entrenament:** L'optimització de l'escalat compost pot requerir un ajustament curós i un coneixement detallat de les capacitats de maquinari i els requisits de l'aplicació.

2.10. CNNs al sector sanitari

Introducció

Les CNNs han revolucionat el camp de l'aprenentatge automàtic i la visió per ordinador i han demostrat un rendiment excel·lent en tasques de classificació d'imatges, detecció d'objectes i segmentació. Al sector sanitari, aquestes capacitats han obert noves possibilitats per millorar el diagnòstic, tractament i seguiment de malalties, oferint eines precises i automatitzades que poden complementar el treball dels professionals de la salut.

Diagnòstic per Imatge

Un dels usos més destacats de les CNNs al sector sanitari és en el diagnòstic per imatge, ja que aquestes poden analitzar imatges mèdiques com ara radiografies, tomografies computeritzades (CT), ressonàncies magnètiques (MRI) i mamografies, detectant anomalies amb una precisió que en alguns casos supera la dels especialistes humans (Litjens et al., 2017) [\[52\]](#).

- **Detecció de Càncer:** Les CNN s'han utilitzat per a la detecció primerenca de càncers com el de mama, pulmó i pell. Per exemple, un model de CNN entrenat en grans conjunts de dades de mamografies pot identificar micro-calcificacions i masses sospitoses, facilitant la detecció primerenca del càncer de mama (Waleed Salehi et al., 2023) [\[53\]](#).

Seguiment de malalties cròniques

Les CNN també tenen un paper crucial en el seguiment de malalties cròniques. Aquestes xarxes poden analitzar seqüències d'imatges al llarg del temps per monitoritzar la progressió de malalties i avaluar l'efectivitat dels tractaments.

- **Retinopatia Diabètica:** Les CNN es fan servir per analitzar imatges de retina i detectar signes de retinopatia diabètica, una complicació de la diabetis que pot portar a la ceguesa si no es tracta adequadament. Els sistemes basats en CNN poden proporcionar diagnòstics ràpids i precisos, permetent intervencions primerenques (Gulshan et al., 2016) [\[54\]](#).

Cirurgia Assistida per Imatge

Les CNN també estan transformant la cirurgia assistida per imatge, on es fan servir per millorar la precisió i la seguretat dels procediments quirúrgics.

- **Segmentació d'Òrgans:** Durant la planificació i l'execució de cirurgies, és crucial identificar i segmentar amb precisió els òrgans i les estructures anatòmiques. Les CNN poden segmentar automàticament òrgans en imatges mèdiques, proporcionant mapes detallats que els cirurgians poden fer servir per guiar els seus procediments (Ronneberger et al., 2015) [\[55\]](#).

Desafiaments i Consideracions

Tot i els seus avantatges, l'ús de CNN al sector sanitari enfronta diversos desafiaments:

- **Qualitat i Quantitat de Dades:** l'entrenament efectiu de les CNNs requereix grans volums de dades etiquetades d'alta qualitat, que poden ser difícils d'obtenir a causa de la privadesa dels pacients i la variabilitat de les imatges mèdiques.
- **Interpretabilitat:** les CNNs són sovint vistes com a "caixes negres" a causa de la complexitat de les seves operacions internes. És crucial desenvolupar mètodes per interpretar i explicar les decisions d'aquestes xarxes per guanyar la confiança dels professionals de la salut.
- **Regulació i Ètica:** l'ús d'intel·ligència artificial a la medicina està subjecte a regulacions estrictes per garantir la seguretat i l'eficàcia. És fonamental abordar les consideracions ètiques relacionades amb l'ús de dades de pacients i la presa de decisions automatitzades (Karimian, Petelos & Evers, 2022) [\[56\]](#).

2.11. CNN a l'anàlisi d'imatges de raigs X

Introducció

Com ja hem comentat a l'apartat anterior, en l'àmbit mèdic, les CNNs han revolucionat l'anàlisi d'imatges mèdiques. En aquest apartat ens centrarem en el seu us pel diagnòstic per raigs X. Aquestes xarxes són especialment eficaces en la detecció de patologies pulmonars, proporcionant una eina poderosa i automatitzada que complementa la tasca dels radiòlegs i millora la precisió del diagnòstic.

Diagnòstic de Patologies Pulmonars

Les malalties pulmonars, com ara la pneumònia, la tuberculosi i, més recentment, la COVID-19, representen un desafiament significatiu per al diagnòstic mèdic. Els raigs X de tòrax són una de les eines més comunes per identificar aquestes malalties. Les CNN han demostrat una notable capacitat per analitzar aquestes imatges i detectar anomalies amb una alta precisió.

- **Pneumònia:** Les CNN poden distingir amb alta precisió entre raigs X normals i aquells que presenten signes de pneumònia. Estudis han mostrat que els models de CNN entrenats en grans conjunts de dades de raigs X poden identificar correctament els patrons d'opacitat i de consolidació típics de la pneumònia (Rajpurkar et al., 2017) [\[57\]](#).
- **Tuberculosi:** La detecció de tuberculosi a través de raigs X pot ser complexa degut a la variabilitat de les presentacions clíniques. Les CNN poden ajudar a identificar signes de tuberculosi, com ara cavitacions i nòduls, millorant la taxa de detecció en comparació dels mètodes tradicionals (Lakhani i Sundaram, 2017) [\[58\]](#).
- **Diagnòstic de Malalties Pulmonars:** Durant la pandèmia de COVID-19, les CNN es van aplicar per analitzar radiografies de tòrax i tomografies computeritzades, ajudant a detectar la presència del virus i avaluar la gravetat de la infecció pulmonar (Shi et al., 2020) [\[59\]](#).

Avantatges de l'Ús de CNN en l'anàlisi Raigs X Toràcics

- **Precisió Diagnòstica:** Les CNN han demostrat ser tan precises, o fins i tot més, que els radiòlegs experimentats en la detecció de certes patologies pulmonars. Això és degut a la seva capacitat per aprendre i reconèixer patrons complexos en les imatges.
- **Velocitat d'Anàlisi:** Un cop entrenades, les CNN poden analitzar raigs X en fraccions de segon, cosa que és crucial en situacions d'emergència on es requereix un diagnòstic ràpid.
- **Escalabilitat:** Els models de CNN es poden desplegar en una varietat d'entorns, des d'hospitals grans fins a clíniques rurals amb accés limitat a radiòlegs especialitzats. Això democratitza l'accés a diagnòstics d'alta qualitat.

Desafiaments i Consideracions

- **Qualitat i quantitat de dades:** L'entrenament efectiu de les CNN requereix grans volums de dades etiquetades d'alta qualitat. La variabilitat en la qualitat de les imatges de raigs X i les diferències en l'anotació de les dades poden afectar el rendiment del model.
- **Interpretabilitat:** Les CNN són sovint vistes com a "caixes negres". És essencial desenvolupar mètodes que permetin interpretar i explicar les decisions d'aquests models per guanyar la confiança dels professionals de la salut i facilitar-ne la integració a la pràctica clínica.
- **Integració Clínica:** La implementació de CNN a l'entorn clínic ha de ser acuradament planificada, assegurant que els models s'integrin de manera fluida amb els fluxos de treball existents i compleixin les normatives i estàndards de salut.

3. Descripció detallada del treball pràctic

3.1. Entorn utilitzat

Tot i que, inicialment, vaig realitzar els primers entrenaments dels diferents models a *Google Colab*, de seguida vaig topiar amb les limitacions, imposades per les seves últimes actualitzacions, de l'ús de la GPU, element que resulta crucial per poder realitzar aquests processos d'una manera ràpid i eficient. Així doncs, després de fer una cerca sobre sistemes alternatius, vaig decidir triar *Kaggle* com a entorn d'execució. Això va suposar una sèrie d'avantatges pel que fa a *Google Colab*: més temps d'execució de GPU, totalment gratuït, una política transparent amb l'ús de recursos i un sistema eficient a l'hora d'importar diferents bases de dades disponibles a la plataforma.

3.2. Base de dades emprada

Base de Dades de Radiografies Toràciques del NIH

La base de dades de radiografies de tòrax de l'Institut Nacional de Salut (NIH) [60] és un conjunt de dades (en endavant *dataset*, per la seva terminologia en anglès) extens i públic que conté imatges de radiografies de tòrax juntament amb anotacions clíniques. Aquest *dataset* inclou més de 100,000 imatges de raigs X de tòrax obtingudes de més de 30,000 pacients, amb 14 tipus de condicions patològiques etiquetades, com ara pneumònia, nòduls i atelèctasi. Aquestes imatges i les etiquetes corresponents són una valuosa eina per a la recerca i el desenvolupament d'algorismes d'aprenentatge automàtic en l'àmbit de la radiologia i la detecció automatitzada de malalties pulmonars (Wang et al., 2017) [64].

Determinació d'Etiquetes a la Base de Dades de Radiografies de Tòrax del NIH

A la base de dades de radiografies de tòrax del NIH, les etiquetes de les imatges es van determinar mitjançant tècniques avançades de processament de llenguatge natural (NLP). Els informes radiològics associats a cada imatge van ser analitzats automàticament per identificar i classificar les condicions patològiques descrites. Aquest procés va permetre extreure 14 tipus de patologies comunes amb una precisió estimada superior al 90%. Tot i això, a causa de la naturalesa automatitzada del procés, és possible que algunes etiquetes siguin errònies. Tot i aquesta limitació, la metodologia emprada proporciona un conjunt de dades extens i representatiu, cosa que facilita l'entrenament de models d'aprenentatge profund per millorar la detecció i el diagnòstic assistit per ordinador en radiografies de tòrax.

Problemes amb l'etiquetatge automàtic de les imatges

Després de fer un parell de sessions d'entrenament, vaig topiar-me amb el fet que els models que estava obtenint tenien una precisió que a penes s'acostava al 70%, fins i tot després de diverses modificacions. Després d'investigar una mica, vaig trobar-me amb el comentari d'un usuari de Kaggle [61], que va topiar amb el mateix problema. En aquest comentari, citava una font com una sèrie d'observacions interessants pel que fa a aquest conjunt de dades [62].

L'article "*Exploring the ChestXray14 dataset: problemes*", per Lauren Oakden-Rayner, identifica diversos problemes amb el *dataset* ChestXray14 utilitzat per entrenar sistemes d'IA mèdica. Els principals problemes inclouen la inexactitud de les etiquetes, la manca de revisió visual de les imatges i la utilitat mèdica limitada de les etiquetes. L'autora mostra que les etiquetes no reflecteixen amb precisió les condicions mèdiques que haurien de representar, cosa que podria portar models d'IA a aprendre patrons incorrectes i fer prediccions mèdicament insensates. L'article conclou que és crucial revisar visualment les imatges per assegurar la validesa de les dades.

Obtenció d'un millor etiquetatge

Al mateix comentari de *Kaggle*, es proporciona un enllaç on s'explica com s'ha millorat l'etiquetatge d'una sèrie d'imatges del *dataset* original [\[63\]](#).

Veiem com el primer estudi es centra en identificar i etiquetar quatre troballes específiques a les radiografies de tòrax: opacitat de l'espai aeri, pneumotòrax, nòdul/massa i fractura. El procés va involucrar la revisió independent de cada imatge per tres radiòlegs certificats. Per a les imatges del conjunt de proves, els radiòlegs van ser seleccionats a l'atzar d'una cohort d'11 radiòlegs certificats. Per a la validació, van participar-hi radiòlegs i residents d'una cohort de 13 individus. Es van seguir aquests passos:

- Revisió Independent: Cada radiòleg va avaluar la presència o absència de les troballes.
- Consens: Si tots tres radiòlegs coincidien, l'etiqueta es confirmava.
- Iteracions: En cas de discrepàncies, es van fer revisions addicionals fins a assolir un consens o es va utilitzar una votació majoritària.

Les etiquetes finals es van registrar en fitxers CSV específics:

- `individual_readers.csv`: Conté les etiquetes individuals de cada radiòleg per a cada imatge.
- `validation_labels.csv` i `test_labels.csv`: Inclouen les etiquetes finals atribuïdes per a les quatre condicions.
- En total, es van etiquetar 4,374 imatges, amb 2,412 per a validació i 1,962 per a proves (Majkowska et al., 2019) [\[65\]](#).

Aquest procés millora significativament la precisió i la utilitat del *data set* NIH ChestX-ray14 per a l'entrenament de models d'aprenentatge profund.

3.3. Pre-processament de dades

Un cop determinat el *dataset*, així com millorada la qualitat de les dades del mateix, es va procedir a la creació d'un *dataset* d'entrenament, classificant les imatges en diferents directoris per a l'entrenament (*train*), validació (*val*) i *testing* (*test*). Així mateix, es van classificar les imatges dins de cadascun d'aquests directoris entre imatges amb i sense la patologia a estudiar. Per això, es van copiar les imatges pertinents del *dataset* "NIH Chest X-rays", usant els fitxers obtinguts en el pas anterior com a referència. Finalment, es va obtenir una distribució com la mostrada a la següent imatge:

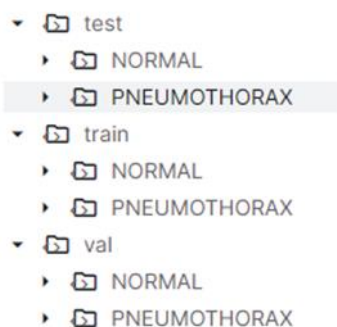


Figura 13: Estructura de les dades d'entrenament

Adicionalment, es va procedir a verificar que les imatges estaven degudament distribuïdes entre els corresponents directoris:

```
Total percentage of images in train directory: 69.99%, Number of images: 2108
Total percentage of images in validation directory: 19.99%, Number of images: 602
Total percentage of images in test directory: 10.03%, Number of images: 302
```

Figura 14: Verificació de la distribució d'imatges d'entrenament

Tots aquests processos van quedar documentats en un quadern de *Kaggle* [66].

Finalment, es va generar el nou *dataset* millorat [67].

3.4. Entrenament dels diferents models

L'entrenament de models de xarxes neuronals convolucionals (CNN) és un procés fonamental en la creació de sistemes de classificació d'imatges. A continuació, es descriuen els passos presos per entrenar diferents models, incloent-hi la importació de biblioteques, la configuració de paràmetres i l'ús de generadores de dades.

Importació de Biblioteques i Mòduls

Per començar, cal importar les biblioteques i els mòduls essencials per a l'entrenament dels models. En aquest cas, s'utilitza *TensorFlow* i *Keras*, juntament amb altres paquets necessaris per a la visualització i l'avaluació de l'exercici del model.

- *TensorFlow* és una biblioteca de codi obert desenvolupada per Google per a la computació numèrica i l'aprenentatge automàtic. La seva escalabilitat permet executar models en diverses plataformes i suporta CPUs, GPUs i TPUs, fet que facilita la construcció, entrenament i avaluació de models de manera eficient. A més, compta amb un ampli ecosistema que inclou diverses eines i una gran comunitat d'usuaris que proporciona recursos i suport abundants.
- *Keras* és una biblioteca d'alt nivell que funciona sobre *TensorFlow*, oferint una interfície simple i amigable per a la construcció i entrenament de models de CNNs. La seva facilitat d'ús permet la creació ràpida de prototips, requisit essencial per a la recerca i desenvolupament en aprenentatge automàtic. S'integra perfectament amb *TensorFlow*, aprofitant-ne la potència i optimització, i és prou versàtil per a la creació de models complexos i personalitzats.

Configuració de Paràmetres

S'estableixen diversos paràmetres importants per al procés d'entrenament, com la mida de les imatges, la mida del lot, el nombre d'èpoques i una llavor per assegurar-ne la reproductibilitat.

- Mida de la Imatge (IMG_SIZE): determina les dimensions a què es redimensionen totes les imatges abans de ser introduïdes al model. És crucial per garantir la consistència en el processament de les imatges, permetent que el model tracti totes les entrades de manera uniforme. A més, la mida de la imatge influeix directament en la complexitat computacional: imatges més grans poden capturar més detalls útils per a la classificació, però també requereixen més memòria i poder de processament. Escollir la mida adequada pot millorar significativament la precisió del model, balancejant la captura de detalls i l'eficiència del model.
- Mida del Lot (BATCH_SIZE): especifica quants exemples d'entrenament es processen junts abans d'actualitzar els paràmetres del model. Lots més grans permeten un ús més eficient dels recursos computacionals, especialment en GPUs, accelerant el procés d'entrenament. A més, una mida de lot adequada ajuda a estabilitzar les actualitzacions dels gradients, millorant la convergència del model. Tot i això, és important equilibrar aquest paràmetre, ja que lots massa grans poden excedir la capacitat de memòria, mentre que lots molt petits poden fer l'entrenament menys eficient i més sorollós.
- Nombre de *Epochs* (EPOCHS): indica quantes vegades passarà tot el conjunt de dades d'entrenament a través del model. És vital per permetre que el model ajusti els pesos adequadament. Un nombre suficient assegura que el model aprengui les característiques de les dades, però un nombre excessiu pot portar al sobre-ajustament on el model s'ajusta massa a les dades d'entrenament i perd capacitat de generalització. A més, més *Epochs* augmenten el temps d'entrenament, per la qual cosa és essencial trobar un equilibri que maximitzi el rendiment sense incórrer en costos computacionals innecessaris.
- Llavor (SEED): s'utilitza per inicialitzar els generadors aleatoris de manera consistent, cosa que és essencial per a la reproductibilitat dels resultats. En fixar una llavor, s'assegura que les transformacions aleatòries aplicades durant l'augment de dades siguin les mateixes en cada execució, cosa que és crucial per fer comparacions justes entre diferents models o configuracions d'híper-paràmetres. Això permet als investigadors i desenvolupadors replicar experiments i validar els resultats de manera fiable, cosa que és fonamental en entorns de recerca i desenvolupament d'aprenentatge automàtic.

Preparació de les Dades

Es defineixen les rutes per als directoris d'entrenament, validació i prova. Després, es crea una instància de *ImageDataGenerator* per augmentar les dades d'entrenament amb diverses transformacions.

```
# Create an instance of ImageDataGenerator with enhanced data augmentation for training data
train_datagen = ImageDataGenerator(
    rescale=1./255, # Rescale pixel values from [0, 255] to [0, 1]
    zoom_range=0.2, # Randomly zoom into images
    rotation_range=20, # Randomly rotate images
    width_shift_range=0.2, # Randomly shift images horizontally
    height_shift_range=0.2, # Randomly shift images vertically
    shear_range=0.2, # Shear intensity
    brightness_range=[0.8, 1.2], # Randomly change brightness
    horizontal_flip=True # Randomly flip images horizontally
)
```

Figura 15: Generador de dades d'imatge

- *rescale=1./255*: escala els valors dels píxels de les imatges d'un rang de [0, 255] a un rang de [0, 1]. Aquest re-escalat és important perquè els models de xarxes neuronals funcionen millor quan les dades d'entrada estan normalitzades. En dividir per 255, s'assegura que els valors dels píxels estiguin dins un rang petit i manejable, millorant l'estabilitat i l'eficiència d'entrenament del model.
- *zoom_range=0.2*: aplica un zoom aleatori a les imatges dins un rang de fins a un 20%. Això ajuda al model a ser més robust davant de variacions a l'escala dels objectes dins de les imatges, fet que pot millorar la seva capacitat de generalització quan s'enfronta a imatges noves que poden tenir objectes de diferents mides.
- *rotation_range=20*: rota les imatges aleatòriament fins a 20 graus en qualsevol direcció, ajudant al model a reconèixer objectes independentment de la seva orientació, cosa que és útil en escenaris on els objectes poden estar orientats de diferents maneres a les imatges de prova.
- *width_shift_range=0.2* i *height_shift_range=0.2*: desplacen les imatges horitzontal i verticalment fins a un 20% de la seva amplada i alçada, respectivament, permetent que el model aprengui a identificar objectes fins i tot quan no estan perfectament centrats en la imatge, augmentant així la robustesa del model davant de variacions en la posició dels objectes.
- *shear_range=0.2*: aplica un cisallament a les imatges amb una intensitat de fins a un 20%, alterant la forma dels objectes a les imatges, ajudant al model a aprendre a identificar objectes fins i tot quan han estat distorsionats de manera no lineal, millorant així la seva capacitat de generalització.
- *brightness_range=[0.8, 1.2]*: ajusta la brillantor de les imatges aleatòriament dins del rang especificat (entre el 80% i el 120% de la brillantor original), permetent al model ser més robust davant de variacions en les condicions d'il·luminació, cosa que és crucial per al reconeixement d'objectes en diferents ambients lumínics.
- *horizontal_flip=True*: inverteix les imatges horitzontalment de manera aleatòria, ajudant al model a aprendre a reconèixer objectes de manera simètrica, millorant-ne el rendiment en escenaris on els objectes poden aparèixer en diferents orientacions laterals.

Per a les dades de validació i prova, s'utilitza una instància de `ImageDataGenerator` que només fa el re-escalat de les imatges. A diferència del conjunt d'entrenament, on s'apliquen diverses tècniques d'augment de dades per millorar la robustesa del model, les dades de validació i prova han de reflectir les condicions reals i no es poden modificar artificialment.

Càrrega de les Dades

Es carreguen les dades de les imatges en lots utilitzant els generadors de dades prèviament definits.

```
train_generator = train_datagen.flow_from_directory(  
    train_dir,  
    target_size=(IMG_SIZE, IMG_SIZE), # Resize images to the specified size  
    class_mode='binary', # Binary classification  
    batch_size=BATCH_SIZE,  
    seed=SEED  
)
```

Figura 16: Càrrega de dades al generador d'entrenament

Definició i Compilació dels Models

En aquest pas, es defineixen i compilen tres models diferents: EfficientNetB0, VGG19 i InceptionV3. Cada model es configura usant una arquitectura pre-entrenada i s'adapta amb capes addicionals per fer una classificació binària. A continuació, s'explica cada pas detalladament:

```
# Function to create EfficientNetB0 model
def create_model_effNet():
    model = Sequential() # Initialize a sequential model
    model.add(EfficientNetB0(include_top=False, weights='imagenet', input_shape=(IMG_SIZE, IMG_SIZE, 3))) # Add EfficientNetB0 base model
    model.add(layers.BatchNormalization()) # Add batch normalization for stabilization
    model.add(layers.GlobalAveragePooling2D()) # Add global average pooling layer
    model.add(layers.Dropout(0.5)) # Add dropout layer for regularization to prevent overfitting
    model.add(layers.Dense(1, activation='sigmoid', kernel_regularizer=l2(0.01))) # Add dense layer for binary classification
    model.compile(loss='binary_crossentropy', optimizer=Adam(learning_rate=1e-4), metrics=['binary_accuracy']) # Compile the model
    return model
```

Figura 17: Definició i compilació d'EfficientNetB0

1. Inicialització del Model Seqüencial: Es crea un model seqüencial, que és una forma d'apilar capes en seqüència.
2. Afegir Model Base EfficientNetB0: S'inclou el model pre-entrenat *EfficientNetB0* sense les capes superiors (*include_top=False*), usant els pesos pre-entrenats de *ImageNet*. Això proporciona una base robusta que ha après a extreure característiques útils de les imatges.
3. Normalització per Lots: La capa de normalització per lots ajuda a estabilitzar i accelerar el procés d'entrenament normalitzant les activacions de les capes anteriors.
4. *Pooling* Global: La capa de *pooling* global redueix les dimensions de les característiques extretes, cosa que ajuda a evitar el sobre-ajustament i redueix la quantitat de paràmetres.
5. *Dropout*: S'hi afegeix una capa de *dropout* amb una taxa del 50% per prevenir el sobre-ajustament, desconnectant aleatòriament la meitat de les neurones durant l'entrenament.
6. Capa Densa: S'hi afegeix una capa densa amb una activació sigmoide per a la classificació binària, juntament amb una regularització L2 per evitar el sobre-ajustament.
7. Compilació del Model: Es compila el model utilitzant la pèrdua binària d'entropia creuada (*binary_crossentropy*) i l'optimitzador Adam amb una taxa d'aprenentatge de 1×10^{-4} .

Tant pel model de VGG19 com per InceptionV3 es segueix un procediment similar.

Entrenament dels Models

El procés d'entrenament dels models en aquest projecte es realitza mitjançant la funció *train_model*, la qual implementa diversos mecanismes per optimitzar l'entrenament i assegurar que els models generin bons resultats. A continuació, es descriuen els més importants:

```
# ModelCheckpoint: Save the best model during training.
model_save = ModelCheckpoint(
    filepath=f'/kaggle/working/best_{model_name}.keras', # Filepath to save the best model
    save_best_only=True, # Save only the model with the best performance on the validation set
    monitor='val_loss', # Metric to monitor for determining the best model
    mode='min', # The mode is set to 'min' because we want to minimize the validation loss
    verbose=1 # Verbose mode to provide detailed logs during model saving
)
```

Figura 18: Codi del Model Checkpoint

Guarda el millor model durant l'entrenament basant-se en el rendiment al conjunt de validació. És particularment útil per evitar el sobre-ajustament, assegurant que s'emmagatzemi el model que millor generalitza les dades de validació.

```
# EarlyStopping: Stop training early if the model performance stops improving.
early_stop = EarlyStopping(
    monitor='val_loss', # Metric to monitor for early stopping
    min_delta=0.001, # Minimum change in the monitored metric to qualify as an improvement
    patience=5, # Number of epochs with no improvement after which training will be stopped
    mode='min', # The mode is set to 'min' because we want to minimize the validation loss
    verbose=1, # Verbose mode to provide detailed logs during early stopping
    restore_best_weights=True # Restore the model weights from the epoch with the best value of the monitored metric
)
```

Figura 19: Codi del Early Stopping

Atura l'entrenament anticipadament si el rendiment del model deixa de millorar. Això prevé el sobre-entrenament i estalvia recursos computacionals en aturar el procés quan les millores addicionals són insignificants.

```
# ReduceLRonPlateau: Reduce the learning rate when the performance metric stops improving.
reduce_lr = ReduceLRonPlateau(
    monitor='val_loss', # Metric to monitor for reducing learning rate
    factor=0.5, # Factor by which the learning rate will be reduced (new_lr = lr * factor)
    patience=3, # Number of epochs with no improvement after which the learning rate will be reduced
    min_delta=0.001, # Minimum change in the monitored metric to qualify as an improvement
    mode='min', # The mode is set to 'min' because we want to minimize the validation loss
    verbose=1 # Verbose mode to provide detailed logs during learning rate reduction
)
```

Figura 20: Codi de reducció del Learning Rate

Redueix la taxa d'aprenentatge quan el rendiment del model deixa de millorar, permetent que el model faci ajustaments més fins, millorant potencialment el rendiment al conjunt de validació.

El procés d'entrenament comença amb la inicialització dels models i la definició de les capes, seguit de la compilació amb els paràmetres adequats com la pèrdua, l'optimitzador i les mètriques. Durant l'entrenament, els models es fitxen utilitzant els generadors de dades d'entrenament i validació, aplicant els *callbacks* esmentats per monitoritzar i ajustar l'entrenament de manera dinàmica. Això assegura que els models aprenguin eficientment sense sobre-ajustar-se, podent generalitzar adequadament les dades noves no vistes. Finalment, els models entrenats es guarden per a la seva posterior utilització i anàlisi.

4. Descripció y anàlisi dels resultats

En aquesta secció, s'avaluaran i es compararan els models entrenats per determinar quin proporciona els millors resultats en la tasca de classificació d'imatges de pneumotòrax. Això es farà a través de diverses mètriques i mètodes d'anàlisi. El primer pas és avaluar els models al conjunt de prova per obtenir prediccions i comparar-les amb les etiquetes veritables.

Anàlisi de la Precisió de l'Entrenament y Validació dels Models

```
# Function to evaluate the model
def evaluate_model(model, model_name):
    test_steps = test_generator.n // test_generator.batch_size
    y_pred = model.predict(test_generator, steps=test_steps, verbose=1)
    y_pred_classes = np.where(y_pred > 0.5, 1, 0) # Convert probabilities to class labels
    y_true = test_generator.classes[:len(y_pred_classes)] # True labels
    return y_true, y_pred, y_pred_classes
```

Figura 21: Funció d'avaluació dels models

Aquesta funció pren com a entrada un model i el seu nom, i realitza l'avaluació al conjunt de prova. Els passos detallats són:

- *test_steps*: Es calcula el nombre de passos necessaris per avaluar tot el conjunt de prova.
- *y_pred*: S'obtenen les prediccions del model al conjunt de prova.
- *y_pred_classes*: Es converteixen les probabilitats predites en etiquetes de classe binàries (0 o 1) usant un llindar de 0.5.
- *y_true*: S'obtenen les etiquetes veritables corresponents a les prediccions generades.

Gràfica de la Pèrdua d'Entrenament i Validació

La gràfica de la pèrdua d'entrenament i la validació proporciona informació valuosa sobre l'aprenentatge dels models durant l'entrenament. La pèrdua (*loss*) mesura la discrepància entre les prediccions del model i les etiquetes de validació. Avaluar tant la pèrdua d'entrenament com la de validació és crucial per entendre si el model està sobre-ajustant (*overfitting*) o infra-ajustant (*underfitting*).

Si la pèrdua de validació és molt més gran que la pèrdua d'entrenament, el model podria estar sobre-ajustant, aprenent massa bé els detalls i el soroll del conjunt d'entrenament i no generalitzant bé en dades noves. Si les dues pèrdues es mantenen altes, el model podria estar infra-ajustat, indicant que no n'ha après prou de les dades d'entrenament. Un model amb pèrdua de validació baixa i estable és preferible, ja que indica una bona capacitat de generalització i un entrenament estable.

A continuació, podem observar les dues gràfiques mencionades, mostrant la comparació d'aquests paràmetres pels models EfficientNetB0 (blau), VGG19 (taronja) i InceptionV3 (verd):

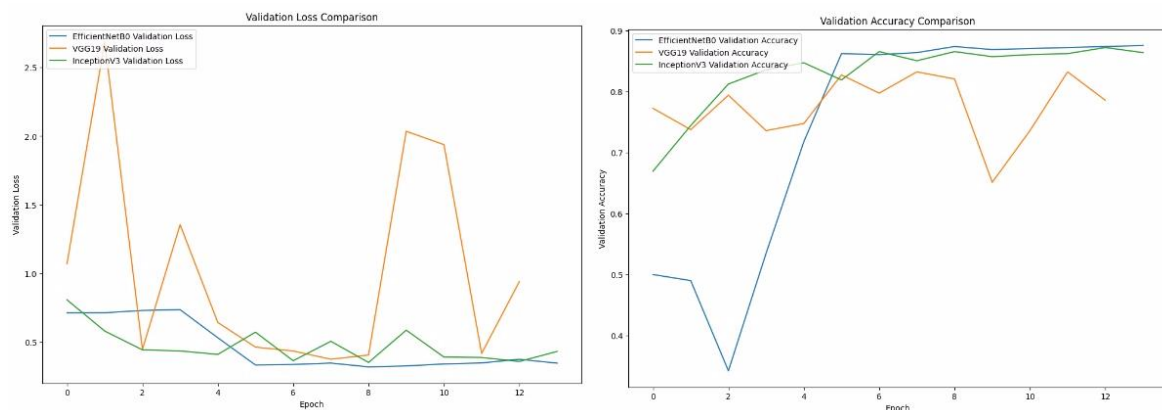


Figura 22: Comparació de la precisió i pèrdua de validació per EfficientNetB0, VGG19 i InceptionV3

Observacions:

- EfficientNetB0: mostra una tendència de pèrdua de validació que disminueix constantment després d'un parell de *Epochs* inicials, on la pèrdua és alta. Això suggereix que, encara que inicialment el model podria haver estat sobre-ajustant o no aprenent correctament, va aconseguir estabilitzar-se i millorar amb el temps.
- VGG19: la seva pèrdua de validació mostra una alta volatilitat amb pics i caigudes significatives, podent indicar que el model està tenint problemes per generalitzar correctament les dades de validació, possiblement degut a sobre-ajustament en alguns *Epochs*.
- InceptionV3: sembla tenir una pèrdua de validació més estable en comparació amb els altres models, mantenint-la consistentment baixa després dels primers *Epochs*, suggerint que aquest està generalitzant bé i no mostrant signes significatius de sobre-ajustament.

Dels tres models, InceptionV3 presenta el millor comportament a la gràfica de pèrdua d'entrenament i validació. Mostra una pèrdua baixa i estable en la validació, cosa que indica una bona capacitat de generalització i un entrenament estable. En comparació, EfficientNetB0 també mostra una millora amb el temps, però VGG19 té problemes d'estabilitat, cosa que suggereix un rendiment menys fiable.

Matriu de Confusió

La matriu de confusió és una eina essencial per avaluar el rendiment del model de classificació. Representa les prediccions del model a una taula que mostra el nombre de prediccions correctes i incorrectes, classificades per cada classe. Cada cel·la indica el nombre d'instàncies predites com a pertanyents a una classe específica en comparació amb les instàncies reals d'aquesta classe.

Aquesta eina és crucial perquè permet identificar com el model confon les diferents classes, proporcionant una visió clara dels tipus específics d'errors que comet. A més, ajuda a avaluar si el model té un biaix cap a una classe particular, assegurant prediccions justes i equilibrades. També facilita el càlcul de mètriques derivades com a precisió, *recall* i F1-score, essencials per avaluar la qualitat de les prediccions del model.

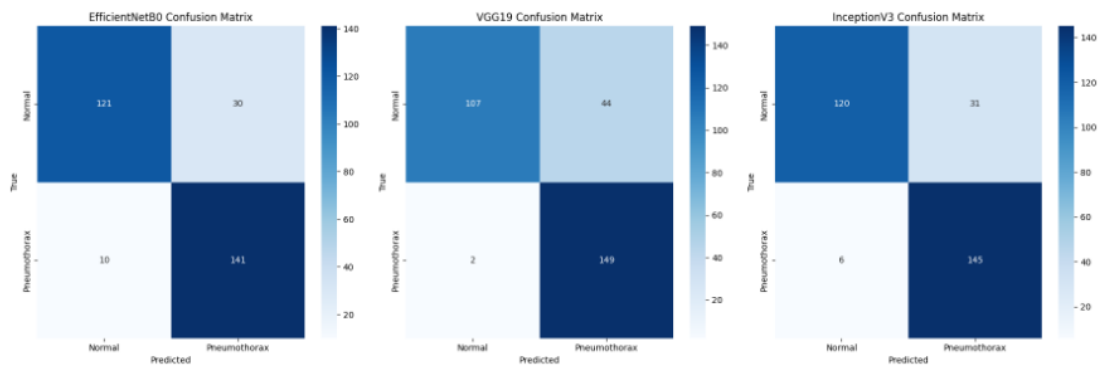


Figura 23: Matriu de confusió de EfficientNetB0, VGG19 i InceptionV3

Observacions:

- EfficientNetB0: mostra 121 prediccions correctes per a "Normal" i 141 per a "Pneumothorax". Hi ha 30 falsos positius i 10 falsos negatius, cosa que indica que el model és força efectiu, però encara té marge de millora en la predicció de la classe "Normal".
- VGG19: mostra 107 prediccions correctes per a "Normal" i 149 per a "Pneumothorax". Tot i això, presenta 44 falsos positius i només 2 falsos negatius. Això suggereix que el model tendeix a predir "Pneumothorax" amb més freqüència, de vegades incorrectament, cosa que pot indicar un biaix cap a la predicció d'aquesta classe.
- InceptionV3: mostra 120 prediccions correctes per a "Normal" i 145 per a "Pneumothorax". Té 31 falsos positius i 6 falsos negatius, cosa que indica un bon equilibri i un rendiment robust similar a EfficientNetB0, però amb una lleugera millora en la precisió per a la classe "Normal".

Dels tres models, tant EfficientNetB0 com InceptionV3 presenten una bona matriu de confusió, aconseguint un equilibri sòlid entre les prediccions correctes de les dues classes i mostrant un nombre reduït de falsos positius i negatius. Això indica un rendiment més robust i equilibrat en comparació amb VGG19.

Corba ROC

La corba ROC (*Receiver Operating Characteristic*) és una representació gràfica que il·lustra el rendiment d'un model de classificació binària mostrant la relació entre la taxa de veritables positius (*True Positive Rate*, TPR) i la taxa de falsos positius (*False Positive Rate*, FPR) en diversos llindars de classificació. L'àrea sota la corba ROC (AUC-ROC) proporciona una mesura agregada del rendiment del model en tots els llindars possibles, on un AUC-ROC més alt indica una capacitat més gran del model per distingir entre les classes.

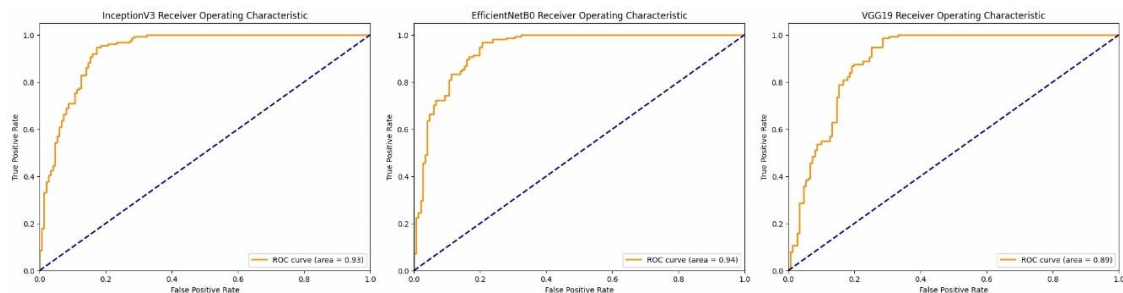


Figura 24: Corba ROC per EfficientNetB0, VGG19 i InceptionV3

Observacions:

- EfficientNetB0: mostra un AUC de 0.94, indicant una excel·lent capacitat per distingir entre les classes "Normal" i "Pneumothorax". La corba s'acosta molt a la cantonada superior esquerra, suggerint una alta taxa de veritables positius i una baixa taxa de falsos positius.
- VGG19: mostra un AUC de 0.89, una mica més baixa comparat amb els altres models. Tot i que segueix sent un bon rendiment, indica que VGG19 té una menor capacitat de discriminació entre les classes comparat amb EfficientNetB0 i InceptionV3.
- InceptionV3: mostra un AUC de 0.93, indicant una capacitat excel·lent per distingir entre les classes. La forma de la corba suggereix un equilibri adequat entre la taxa de veritables positius i la taxa de falsos positius, semblant a EfficientNetB0, però lleugerament inferior en termes d'AUC.

Dels tres models, EfficientNetB0 presenta la millor corba ROC amb un AUC de 0,94, indicant una excel·lent capacitat per distingir entre les classes "Normal" i "Pneumothorax". InceptionV3 també mostra un rendiment molt fort amb un AUC de 0,93. D'altra banda, encara que VGG19 té un bon rendiment, el seu AUC de 0,89 és inferior al dels altres dos models, indicant una capacitat de discriminació lleugerament menor. En general, EfficientNetB0 es destaca com el millor en termes de la mètrica AUC-ROC, però tan sols lleugerament davant de InceptionV3.

Corba de Precisió-Recall

La corba de precisió-recall mostra la relació entre la precisió (precision) i el recall (sensibilitat) a diversos llindars de classificació. L'àrea sota la corba de precisió-recall (PR-AUC) proporciona una mesura agregada del rendiment del model en tots els llindars possibles. Aquesta corba és particularment útil en conjunts de dades desbalancejades, on una alta precisió pot no ser suficient si el recall és baix.

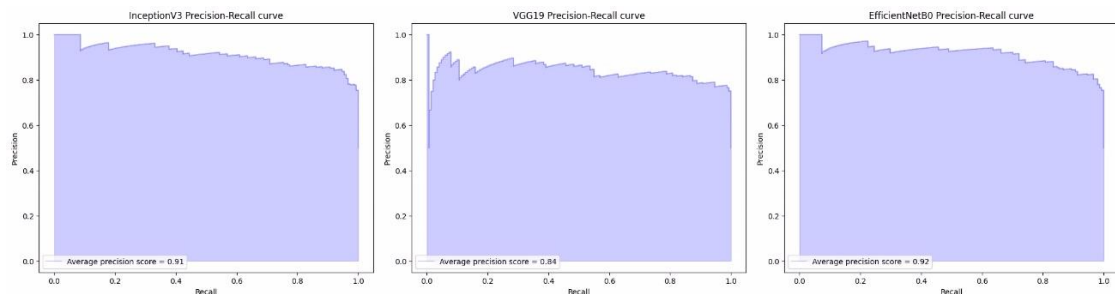


Figura 25: Corba de Precisió-Recall per EfficientNetB0, VGG19 i InceptionV3

Observacions:

- EfficientNetB0: La corba de precisió-recall per a EfficientNetB0 mostra un PR-AUC de 0,92, indicant un excel·lent equilibri entre precisió i recall. Això significa que el model manté una alta precisió mentre captura la majoria dels positius veritables, cosa que és ideal per a tasques de classificació on ambdós aspectes són importants.
- VGG19: La corba de precisió-recall per a VGG19 mostra un PR-AUC de 0,84, la qual cosa suggereix un rendiment inferior en comparació amb els altres models. Encara que la precisió és alta en alguns llindars, el recall no és tan fort, cosa que pot ser un inconvenient en escenaris on és crític identificar tots els positius veritables.
- InceptionV3: La corba de precisió-recall d'InceptionV3 mostra un PR-AUC de 0,91, cosa que també indica un rendiment excel·lent. Similar a EfficientNetB0, aquest model aconsegueix un bon equilibri entre precisió i recall, encara que lleugerament inferior al PR-AUC d'EfficientNetB0.

Dels tres models, EfficientNetB0 presenta la millor corba de precisió-recall amb un PR-AUC de 0,92, demostrant la seva capacitat per mantenir un alt nivell de precisió mentre captura la majoria dels veritables positius. InceptionV3 també mostra un rendiment fort amb un PR-AUC de 0,91. En comparació, VGG19 té un rendiment menor amb un PR-AUC de 0,84, indicant que, encara que pot ser precís, no captura tants positius veritables com els altres models. En resum, EfficientNetB0 es destaca com el millor en termes de la mètrica PR-AUC, seguit de prop per InceptionV3.

Resum i Comparació de Mètriques (Exactitud, Precisió, Recall, F1 Score, AUC-ROC, PR-AUC)

El resum i la comparació de mètriques proporcionen una visió completa del rendiment dels models en diferents aspectes. Les mètriques clau inclouen:

- **Exactitud (*Accuracy*):** Proporció de prediccions correctes sobre el total de prediccions.
- **Precisió (*Precision*):** Proporció de veritables positius sobre el total de positius predits.
- **Recall (Sensibilitat):** Proporció de veritables positius sobre el total de veritables positius i falsos negatius.
- **F1 Score:** Mitjana harmònica de precisió i *recall*.
- **AUC-ROC:** Àrea sota la corba ROC.
- **PR-AUC:** Àrea sota la corba de precisió-*recall*.

Aquestes mètriques juntes proporcionen una avaluació robusta del rendiment del model, ajudant a comparar models de manera objectiva i seleccionar el millor per a la tasca en qüestió.

Anàlisi dels Models:

- **EfficientNetB0:**
 - Exactitud: 0.8675
 - Precisió: 0.8246
 - Recall: 0.9338
 - F1 Score: 0.8758
 - AUC-ROC: 0.9376
 - PR-AUC: 0.9177

EfficientNetB0 mostra un rendiment excel·lent en termes d'AUC-ROC (0.9376) i PR-AUC (0.9177), cosa que indica una capacitat excel·lent de discriminació entre classes i un bon equilibri entre precisió i *recall*.

- **VGG19:**
 - Exactitud: 0.8477
 - Precisió: 0.7720
 - Recall: 0.9868
 - F1 Score: 0.8663
 - AUC-ROC: 0.8934
 - PR-AUC: 0.8405

VGG19 presenta un recall molt alt (0.9868), cosa que significa que identifica gairebé tots els veritables positius. Tot i això, la seva precisió (0.7720) i PR-AUC (0.8405) són inferiors en comparació amb els altres models, la qual cosa suggereix que té més falsos positius.

- **InceptionV3:**
 - Exactitud: 0.8775
 - Precisió: 0.8239
 - Recall: 0.9603
 - F1 Score: 0.8869
 - AUC-ROC: 0.9327
 - PR-AUC: 0.9121

InceptionV3 també mostra un rendiment excel·lent amb una alta exactitud (0.8775) i un bon equilibri entre precisió (0.8239) i recall (0.9603). Els valors d'AUC-ROC (0.9327) i PR-AUC (0.9121) són lleugerament inferiors als d'EfficientNetB0, però segueix sent un model molt robust.

5. Conclusions

En aquest treball s'han comparat tres arquitectures de xarxes neuronals convolucionals (CNN) – EfficientNetB0, VGG19 i InceptionV3– aplicades a la detecció de malalties pulmonars en radiografies toràciques. Tot seguit, es resumeixen les conclusions principals obtingudes a partir d'aquesta investigació.

Rendiment dels Models

Els models EfficientNetB0 i InceptionV3 han demostrat un rendiment superior en comparació amb VGG19, tant en termes de precisió com de capacitat de generalització. EfficientNetB0 ha mostrat una precisió general del 86.75% i una àrea sota la corba ROC (AUC-ROC) de 0.9376, mentre que InceptionV3 ha aconseguit una precisió del 87.75% i un AUC-ROC de 0.9327. VGG19, encara que mostra un alt valor de *recall* (98,68%), presenta una precisió més baixa (77,20%) i un AUC-ROC de 0,8934, indicant una menor capacitat per distingir entre les classes.

Robustesa i Generalització

EfficientNetB0 i InceptionV3 han demostrat ser més robustos i equilibrats en les prediccions. Tots dos models mostren una baixa taxa de falsos positius i falsos negatius, cosa que suggereix una bona capacitat per generalitzar en dades no vistes. D'altra banda, VGG19 presenta una alta variabilitat en la pèrdua de validació, suggerint possibles problemes de sobre-ajustament, així com una menor estabilitat a l'entrenament.

Avantatges i Limitacions

EfficientNetB0 destaca pel seu equilibri òptim entre precisió i eficiència computacional, i és altament efectiu per a la classificació d'imatges mèdiques amb menys recursos computacionals. InceptionV3 també ha mostrat un rendiment robust i estable, encara que la seva complexitat arquitectònica pot ser una limitació en quant a la implementació i l'ajustament.

Aplicacions Futures

La implementació de CNNs a la detecció de malalties pulmonars mitjançant radiografies toràciques té el potencial de millorar significativament el diagnòstic mèdic, oferint una eina precisa i ràpida que complementa el treball dels professionals de la salut. La capacitat d'aquestes xarxes per identificar i classificar anomalies pulmonars amb alta precisió pot contribuir a una detecció primerenca i al tractament més efectiu d'aquestes patologies, millorant així els resultats clínics i la qualitat de vida dels pacients.

Recomanacions per a Investigacions Futures

Aquest estudi estableix les bases per a futures investigacions en el camp de la detecció automatitzada de malalties mitjançant imatges mèdiques. Es recomana explorar la combinació de diferents models i tècniques d'aprenentatge profund, així com la millora contínua dels conjunts de dades utilitzades per a l'entrenament, assegurant la qualitat i la representativitat de les etiquetes. A més, és crucial desenvolupar mètodes per interpretar i explicar les decisions dels models, facilitant la seva integració a la pràctica clínica i guanyant la confiança dels professionals de la salut.

6. Bibliografia

Machine Learning:

- [1] *What is Machine learning?* [en línia] [data de consulta: 7 d'Abril] Disponible a: <https://cloud.google.com/learn/what-is-machine-learning>
- [2] *Aprendizaje automático: Qué es y por qué es importante* [en línia] [data de consulta: 7 d'Abril] Disponible a: https://www.sas.com/es_es/insights/analytics/machine-learning.html
- [3] *Machine Learning: definición, funcionamiento, usos* [en línia] [data de consulta: 7 d'Abril] Disponible a: <https://datascientest.com/es/machine-learning-definicion-funcionamiento-usos>
- [4] *Racial Discrimination in Face Recognition Technology* [en línia] [data de consulta: 10 d'Abril] Disponible a: <https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/>
- [5] Gentzel, M. Biased Face Recognition Technology Used by Government: A Problem for Liberal Democracy. *Philos. Technol.* 34, 1639–1663 (2021). Disponible a: <https://doi.org/10.1007/s13347-021-00478-z>
- [6] *Supervised versus unsupervised learning: What's the difference?* [en línia] [data de consulta: 15 d'Abril] Disponible a: <https://www.ibm.com/think/topics/supervised-vs-unsupervised-learning>
- [7] A Survey and Perspective on Artificial Intelligence for Security-Aware Electronic Design Automation - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/a-A-categorization-of-major-machine-learning-techniques-with-relevant-examples-b-An_fig1_360078910 [accessed 13 Jun, 2024]

Algoritmes de classificació:

- [8] Alnuaimi, Amer & Albaldawi, Tasnim. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences.* 97. 00133. 10.1051/bioconf/20249700133. Disponible a: https://www.researchgate.net/publication/379633049_An_overview_of_machine_learning_classification_techniques
- [9] *Classification algorithms: Definition and main models* [en línia] [data de consulta: 4 de Maig] Disponible a: <https://datascientest.com/en/classification-algorithms-definition-and-main-models>
- [10] *Train Test Validation Split: How To & Best Practices* [2023] [en línia] [data de consulta: 5 de Juny] Disponible en: <https://www.v7labs.com/blog/train-validation-test-set>

Deep Learning:

- [11] *Diferencia entre machine learning y deep learning* [en línia] [data de consulta: 17 de Maig] Disponible a: <https://blogs.uoc.edu/informatica/es/machine-learning-vs-deep-learning-diferencias/>
- [12] *¿Qué es Deep Learning?* [en línia] [data de consulta: 18 d'Abril] Disponible a: <https://www.ibm.com/es-es/topics/deep-learning>
- [13] *Deep Learning: Qué es y por qué es importante* [en línia] [data de consulta: 18 d'Abril] Disponible a: https://www.sas.com/es_es/insights/analytics/deep-learning.html
- [14] *Deep Learning o Aprendizaje profundo : ¿qué es?* [en línia] [data de consulta: 18 d'Abril] Disponible a: <https://datascientest.com/es/deep-learning-definicion>

- [15] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- [16] GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep learning*. MIT press, 2016.
- [17] Neural Network Definition and Components [en línia] [data de consulta: 3 de Juny] Disponible en: <https://s.mriquestions.com/what-is-a-neural-network.html>
- [18] Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Deep-learning-diagram_fig5_323784695 [accessed 13 Jun, 2024]
- [19] Neurons, Activation Functions, Back-Propagation, Epoch, Gradient Descent: What are these? [en línia] [data de consulta: 22 de Juny] Disponible en: <https://towardsdatascience.com/neurons-activation-functions-back-propagation-epoch-gradient-descent-what-are-these-c80349c6c452>
- [20] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.

CNNs:

- [21] Hardesty, Larry. Explained: Neural networks. Ballyhooed artificial-intelligence technique known as "deep learning" revives 70-year-old idea. En: *MIT News* [en línia]. 14 d'Abril del 2017. Disponible a: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>
- [22] *What is a neural network?* [en línia] [data de consulta: 14 d'Abril] Disponible a: <https://www.ibm.com/topics/neural-networks>
- [23] *¿Qué es una red neuronal?* [en línia] [fecha de consulta: 14 d'Abril] Disponible a: <https://aws.amazon.com/es/what-is/neural-network/>
- [24] *¿Cómo funcionan las Convolutional Neural Networks? Visión por Ordenador* [en línia] [data de consulta: 14 d'Abril] Disponible a: <https://www.aprendemachinelearning.com/como-funcionan-las-convolutional-neural-networks-vision-por-ordenador/>
- [25] *Convolutional Neural Network : definición y funcionamiento* [en línia] [data de consulta: 28 d'Abril] Disponible en: <https://datascientest.com/es/convolutional-neural-network-es>
- [26] *¿Qué son las redes neuronales convolucionales? 3 cosas que debes saber* [en línia] [data de consulta: 30 d'Abril] Disponible en: <https://es.mathworks.com/discovery/convolutional-neural-network.html>
- [27] A lightweight CORONA-NET for COVID-19 detection in X-ray images - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Schematic-of-Convolutional-Neural-Network-CNN_fig4_369963460 [accessed 23 Jun, 2024]
- [28] Muwei Jian, Runxia Zhao, Xin Sun, Hanjiang Luo, Wenyin Zhang, Huaxiang Zhang, Junyu Dong, Yilong Yin, Kin-Man Lam, Saliency detection based on background seeds by object proposals and extended random walk, *Journal of Visual Communication and Image Representation*, Volume 57, 2018, Pages 202-211, ISSN 1047-3203, <https://doi.org/10.1016/j.jvcir.2018.11.007>.
- [29] Surek, Guilherme Augusto Silva, Laio Oriel Seman, Stefano Frizzo Stefenon, Viviana Cocco Mariani, and Leandro dos Santos Coelho. 2023. "Video-Based Human Activity Recognition Using Deep Learning Approaches" *Sensors* 23, no. 14: 6384. <https://doi.org/10.3390/s23146384>

Transfer Learning:

- [30] *Introducción a Transfer Learning* [en línia] [data de consulta: 23 de Maig] Disponible a: <https://es.mathworks.com/discovery/transfer-learning.html>
- [31] *¿Qué es el transfer learning y qué ventajas tiene?* [en línia] [data de consulta: 23 de Maig] Disponible a: <https://www.unir.net/ingenieria/revista/transfer-learning/>
- [32] *What is transfer learning?* [en línia] [ata de consulta: 23 de Maig] Disponible a: <https://www.ibm.com/topics/transfer-learning>
- [33] *A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning* [en línia] [data de consulta: 23 de Maig] Disponible a: <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>
- [34] *Transfer Learning: What is it?* [en línia] [data de consulta: 23 de Maig] Disponible a: <https://datascientest.com/en/transfer-learning-what-is-it>
- [35] *Learning Transfer Learning* [en línia] [data de consulta: 27 de Maig] Disponible a: <https://medium.com/metaor-artificial-intelligence/learning-transfer-learning-31b4b05f5a1a>
- [36] *¿Qué es el aprendizaje por transferencia?* [en línia] [data de consulta: 30 de Maig] Disponible en: <https://aws.amazon.com/es/what-is/transfer-learning/>
- [37] *Introducing Transfer Learning as Your Next Engine to Drive Future Innovations* [en línia] [data de consulta: 7 de Juny] Disponible a: <https://medium.datadriveninvestor.com/introducing-transfer-learning-as-your-next-engine-to-drive-future-innovations-5e81a15bb567>

Ús de Models Pre-entrenats a Xarxes Neuronals Convolucionals (CNN):

- [38] *Difference between AlexNet, VGGNet, ResNet, and Inception* [en línia] [data de consulta: 8 d'Abril] Disponible a: <https://towardsdatascience.com/the-w3h-of-alexnet-vggnet-resnet-and-inception-7baaecccc96>
- [39] *A Deep Dive into Pretrained Models: VGG-16, VGG-19, ResNet, AlexNet, and Inception* [en línia] [data de consulta: 4 de Juny] Disponible a: <https://medium.com/@abidaubaid229/a-deep-dive-into-pretrained-models-vgg-16-vgg-19-resnet-alexnet-and-inception-40df8cfa38c1>

InceptionV3

- [40] *Guía avanzada de Inception v3* [en línia] [data de consulta: 30 de Maig] Disponible a: <https://cloud.google.com/tpu/docs/inception-v3-advanced?hl=es-419>
- [41] *Inception V3 CNN Architecture Explained* [en línia] [data de consulta: 30 de Maig] Disponible a: <https://medium.com/@AnasBrital98/inception-v3-cnn-architecture-explained-691cfb7bba08>
- [42] *Inception-v3 Architecture (Batch Norm and ReLU are used after Conv)* [en línia] [data de consulta: 15 de Juny] Disponible a: <https://cloud.google.com/tpu/docs/inception-v3-advanced>
- [43] Szegedy, Christian & Vanhoucke, Vincent & Ioffe, Sergey & Shlens, Jon & Wojna, ZB. (2016). Rethinking the Inception Architecture for Computer Vision. 10.1109/CVPR.2016.308.

VGG19

- [44] VGG Very Deep Convolutional Networks (VGGNet) – What You Need to Know [en línia] [data de consulta: 12 d'Abril] Disponible a: <https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/>
- [45] *Image Detection Using the VGG-19 Convolutional Neural Network* [en línia] [data de consulta: 12 d'Abril] Disponible a: <https://melisabardhi.medium.com/image-detection-using-convolutional-neural-networks-89c9e21ffa3>
- [46] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556.
- [47] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2014). ImageNet Large Scale Visual Recognition Challenge. arXiv preprint arXiv:1409.0575.
- [48] *Transfer Learning for Illustration Classification - Scientific Figure on ResearchGate*. Available from: https://www.researchgate.net/figure/Architecture-of-the-network-VGG19-The-network-has-16-convolutions-with-ReLUs-NH10_fig3_325642953 [accessed 18 Jun, 2024]
- [49] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition.

EfficientNetB0

- [50] Tan, M. and Le, Q.V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, 9-15 June 2019, 6105-6114. <http://proceedings.mlr.press/v97/tan19a.html>
- [51] *Empirical Analysis of a Fine-Tuned Deep Convolutional Model in Classifying and Detecting Malaria Parasites from Blood Smears - Scientific Figure on ResearchGate*. Available from: https://www.researchgate.net/figure/EfficientNetB0-baseline-model-architecture-33_fig2_348915715 [accessed 18 Jun, 2024]

CNNs en el sector Sanitari:

- [52] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, Clara I. Sánchez, A survey on deep learning in medical image analysis, *Medical Image Analysis*, Volume 42, 2017, Pages 60-88, ISSN 1361-8415, <https://doi.org/10.1016/j.media.2017.07.005>.
- [53] Salehi, Ahmad Waleed, Shakir Khan, Gaurav Gupta, Bayan Ibrahim Alabdullah, Abrar Almjally, Hadeel Alsolai, Tamanna Siddiqui, and Adel Mellit. 2023. "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope" *Sustainability* 15, no. 7: 5930. <https://doi.org/10.3390/su15075930>
- [54] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*. 2016 Dec 13;316(22):2402-2410. doi: 10.1001/jama.2016.17216. PMID: 27898976.

- [55] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
- [56] Karimian, G., Petelos, E. & Evers, S.M.A.A. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. *AI Ethics* 2, 539–551 (2022). <https://doi.org/10.1007/s43681-021-00131-7>

CNNs en l'anàlisi de raigs X:

- [57] Rajpurkar, Pranav & Irvin, Jeremy & Zhu, Kaylie & Yang, Brandon & Mehta, Hershel & Duan, Tony & Ding, Daisy & Bagul, Aarti & Langlotz, Curtis & Shpanskaya, Katie & Lungren, Matthew & Ng, Andrew. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.
- [58] Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*. 2017 Aug;284(2):574-582. doi: 10.1148/radiol.2017162326. Epub 2017 Apr 24. PMID: 28436741.
- [59] Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, He K, Shi Y, Shen D. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. *IEEE Rev Biomed Eng*. 2021;14:4-15. doi: 10.1109/RBME.2020.2987975. Epub 2021 Jan 22. PMID: 32305937.

Base de dades emprada:

- [60] NIH Clinical Center. 2017. "NIH Chest X-rays." Kaggle. Accessed June 23, 2024. <https://www.kaggle.com/datasets/nih-chest-xrays/data/data>.
- [61] SquishNChips. "How is this dataset from 2017?" Comment on NIH Chest X-rays, December 23, 2021. <https://www.kaggle.com/datasets/nih-chest-xrays/data/discussion/300917>.
- [62] Oakden-Rayner, Lauren. "Exploring the ChestXray14 Dataset: Problems." *Dr. Lauren Oakden-Rayner (blog)*. December 18, 2017. <https://laurenokdenrayner.com/2017/12/18/the-chestxray14-dataset-problems/>.
- [63] Google Cloud. "NIH Chest X-rays: Additional Labels." Last modified 2023. https://cloud.google.com/healthcare-api/docs/resources/public-datasets/nih-chest#additional_labels.
- [64] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald Summers, ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, *IEEE CVPR*, pp. 3462-3471, 2017
- [65] Anna Majkowska, Sid Mittal, David F. Steiner, Joshua J. Reicher, Scott Mayer McKinney, Gavin E. Duggan, Krish Eswaran, PoHsuan Cameron Chen, Yun Liu, Sreenivasa Raju Kalidindi, Alexander Ding, Greg S. Corrado, Daniel Tse, Shravya Shetty, Chest Radiograph Interpretation Using Deep Learning Models: Assessment Using Radiologist Adjudicated Reference Standards and Population-Adjusted Evaluation, *Radiology*, 2019.

Notebooks de Kaggle

- [66] Simó Hospedales Guillem. *Extract Improved Data*. Kaggle. Disponible a: <https://www.kaggle.com/code/guillemshospedales/extract-improved-data>.
- [67] Simó Hospedales Guillem. *Improved Pneumothorax Training Data*. Kaggle. Disponible a: <https://www.kaggle.com/datasets/guillemshospedales/improved-pneumothorax-training-data/settings>.
- [68] Simó Hospedales Guillem. *Comparing CNN Models*. Disponible a: <https://www.kaggle.com/code/guillemshospedales/comparing-cnn-models/notebook>