

**MurrAI Quality: desenvolupament d'un sistema de *deep learning*
per a l'avaluació automatitzada de la qualitat de mostres d'esput en
microbiologia clínica**

Laia Martos i Massó

TREBALL FINAL DE GRAU BIOTECNOLOGIA

Tutor acadèmic: Miguel Mulero Abellán
Doctor en Ciencias, Llicenciatura en Biologia
Departament de Bioquímica i Biotecnologia
miquel.mulero@urv.cat

En cooperació amb: Centre de Diagnòstic Biomèdic
Hospital Clínic i Provincial de Barcelona

Supervisor/s: Dr. Climent Casals Pascual
Cap del Servei de Microbiologia
Centre de Diagnòstic Biomèdic
ccasals@clinic.cat

Setembre 2025

Jo, Laia Martos i Massó , amb DNI 21779979Z, sóc coneixedora de la guia de prevenció del plagi a la URV Prevenció, detecció i tractament del plagi en la docència: guia per a estudiants (aprovada el juliol 2017)

(<http://www.urv.cat/ca/vidacampus/serveis/crai/que-us-oferim/formaciocompetencies-nuclears/plagi/>) i afirmo que aquest TFG no constitueixen cap de les conductes considerades com a plagi per la URV.

Tarragona, 30 d'agost de 2025

Laia Martos Massó

AGRAÏMENTS

Vull expressar el meu agraïment al Servei de Microbiologia del Centre de Diagnòstic Biomèdic (CDB) de l'Hospital Clínic de Barcelona per acollir-me i oferir-me una molt bona estada durant la realització d'aquest treball. En especial voldria agrair al meu tutor del CDB, el Dr. Climent Casals, per confiar en mi, per donar-me l'oportunitat de realitzar aquest projecte i per la seva coordinació i orientació. També vull agrair al meu tutor acadèmic, Miquel Mulero, pel seu seguiment i per les seves aportacions i correccions al treball.

També vull donar les gràcies al Dr. Mateu Espasa i a l'Andrea Vergara per la seva ajuda durant la meva estada al centre, per actuar com a supervisors i per ensenyar-me amb paciència i dedicació.

D'igual manera m'agradaria agrair a Rubén López Aladid, que m'ha ajudat i guiat en la creació del model d'intel·ligència artificial. La seva ajuda ha estat clau en el desenvolupament d'aquest projecte i sense ell no hauria estat possible realitzar-lo. A més, també m'ha despertat un gran interès en aquest camp i m'ha fet reflexionar sobre el camí que vull seguir en la meva carrera professional.

Finalment, voldria dedicar un agraïment especial a la meva família i a les persones que m'han acompanyat i han confiat en mi tant durant aquests últims 4 anys com durant la realització d'aquest projecte.

ÍNDEX

DADES DEL CENTRE.....	4
RESUM I PARAULES CLAU.....	5
INTRODUCCIÓ.....	6
1. Rellevància clínica de les infeccions respiratòries.....	6
2. Les secrecions respiratòries com a mostra diagnòstica.....	6
3. Qualitat de la mostra i valor diagnòstic.....	7
3.1. Tinció de Gram.....	9
3.2. Criteris d'avaluació.....	10
4. Limitacions dels mètodes convencionals.....	11
5. Intel·ligència artificial.....	11
5.1. Deep learning.....	12
5.2. Xarxes neuronals convolucionals.....	13
5.3. Tasques de visió per computador i aplicació clínica.....	14
5.4. Limitacions i consideracions per la implementació de la IA a l'àmbit mèdic	15
HIPÒTESIS DE TREBALL I OBJECTIUS.....	16
METODOLOGIA.....	17
1. Adquisició de les imatges.....	17
2. Anotació del conjunt de dades.....	18
3. Arquitectura de la CNN.....	20
4. Estratègia d'entrenament de la CNN.....	23
4.1. Augment de dades.....	25
5. Interfície gràfica.....	26
RESULTATS.....	27
DISCUSSIÓ.....	33
CONCLUSIONS.....	36
AUTOAVALUACIÓ.....	37
BIBLIOGRAFIA.....	38

DADES DEL CENTRE

He realitzat les meves pràctiques curriculars i el present Treball de Fi de Grau al Servei de Microbiologia del Centre de Diagnòstic Biomèdic (CDB) de l'Hospital Clínic de Barcelona. El CDB és un laboratori especialitzat en el diagnòstic biomèdic centrat en el pacient, que ofereix serveis de qualitat a través de la innovació tecnològica i la formació contínua dels seus professionals. Compta amb més de 400 col·laboradors, dels quals aproximadament 100 són facultatius de les diferents especialitats del laboratori (Centre de Diagnòstic Biomèdic | Hospital Clínic Barcelona, 2025).

El Servei de Microbiologia és dirigit pel Dr. Climent Casals Pascual, qui també ha actuat com el meu tutor durant la meua estada de pràctiques. El Servei de Microbiologia proporciona informació per a la prevenció, diagnòstic, tractament i seguiment de malalties d'origen infecciosos. Aquest consta de dues seccions principals, Bacteriologia i Virologia, i una Unitat de Parasitologia.

El Servei de Microbiologia compta amb una trajectòria reconeguda en recerca, tant en projectes propis com en col·laboracions amb equips nacionals i internacionals. La seva activitat investigadora es centra principalment en aplicar els coneixements microbiològics a l'àmbit clínic, amb línies de treball que inclouen el desenvolupament de noves tècniques diagnòstiques, l'estudi de la patogènia de les malalties infeccioses, l'anàlisi molecular de la resistència als antimicrobians, la identificació de nous patògens i microorganismes importats i la recerca sobre el microbioma intestinal (Centre de Diagnòstic Biomèdic | Hospital Clínic Barcelona, 2025).

RESUM I PARAULES CLAU

L'avaluació de la qualitat de les mostres d'esput mitjançant la tinció de Gram i un microscopi és un pas essencial per obtenir resultats fiables en el diagnòstic microbiològic d'infeccions respiratòries del tracte respiratori inferior, però la seva interpretació manual pot ser subjectiva i requereix temps i personal qualificat. En aquest treball s'ha desenvolupat un algorisme anomenat MurrAI Quality, que es tracta d'un sistema automatitzat basat en xarxes neuronals convolucionals capaç de detectar i comptar cèl·lules epitelials escamoses i leucòcits en imatges microscòpiques per classificar les mostres segons la seva qualitat. El model s'ha entrenat amb 402 microfotografies d'esput anotades manualment, indicant la localització de leucòcits i cèl·lules epitelials per tal de proporcionar al model exemples dels patrons a reconèixer. Finalment, s'ha avaluat la validesa del sistema en tasques de segmentació d'instàncies i classificació. Aquest estudi constitueix una prova de concepte amb potencial d'implementació en entorns clínics, que obre la porta a futures optimitzacions.

Paraules clau: xarxes neuronals convolucionals, YOLOv8-seg, segmentació, classificació, esput, control de qualitat, microbiologia clínica.

INTRODUCCIÓ

1. Rellevància clínica de les infeccions respiratòries

Les infeccions del tracte respiratori inferior (ITRI) suposen una de les principals causes de morbiditat i mortalitat a nivell mundial, afectant grups de totes les edats. Segons estudis recents, l'any 2021 es van reportar 2,18 milions de morts atribuïbles a ITRIs, sense tenir en compte aquells casos relacionats amb la COVID-19 (Bender et al., 2024). Aquest conjunt de patologies, que inclouen afeccions com la pneumònia, la bronquitis aguda, la grip i la tuberculosi, són principalment causades per bacteris patògens com *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis* i *Pseudomonas aeruginosa*. La detecció precisa d'aquests patògens és fonamental per establir un diagnòstic fiable i per administrar el tractament antimicrobià precoç apropiat que eviti complicacions clíniques greus (Bouzada et al., 2025).

2. Les secrecions respiratòries com a mostra diagnòstica

Tenint en compte la importància d'un diagnòstic acurat, l'esput resulta un element clau d'estudi microbiològic de les infeccions respiratòries, ja que constitueix la mostra principal procedent de les vies respiratòries inferiors. Es tracta d'una secreció composta per una barreja de saliva i moc que prové de les vies respiratòries, habitualment en cas d'infecció o irritació de la mucosa. El seu contingut inclou, a part de microorganismes comensals i patògens, diversos components moleculars com lípids solubles i proteïnes i components cel·lulars com cèl·lules epitelials i leucòcits (Shen et al., 2025).

La recollida de la mostra és, en general, relativament senzilla i poc invasiva. L'opció habitual és obtenir mostres de les vies respiratòries inferiors mitjançant la tos voluntària del pacient. No obstant això, en determinades situacions clíniques pot ser necessari adoptar estratègies més invasives per obtenir mostres diagnòstiques, especialment quan el pacient presenta dificultats per expectorar les secrecions. En aquest cas, es poden utilitzar mètodes d'aspiració traqueal, inducció amb inhalació

salina o succió bronquial a través d'un catèter, especialment en pacients d'UCI intubats (Cartulieres et al., 2023). Addicionalment, altres tècniques auxiliars de fisioteràpia també s'utilitzen com a adjuvant per facilitar l'obtenció de l'espècimen a analitzar (Shen et al., 2025).

Després de la recollida d'esput, s'identifica i selecciona la porció purulenta de la mostra. Tot seguit, s'elabora un frotis i s'aplica la tinció de Gram, que permet una avaluació de la qualitat mitjançant l'anàlisi del nombre de leucòcits i cèl·lules epitelials, així com una primera orientació sobre el possible agent etiològic. Si la mostra compleix els criteris de qualitat establerts, s'inocula en els medis de cultiu corresponents amb l'objectiu d'aïllar i identificar el microorganisme responsable i confirmar un diagnòstic.

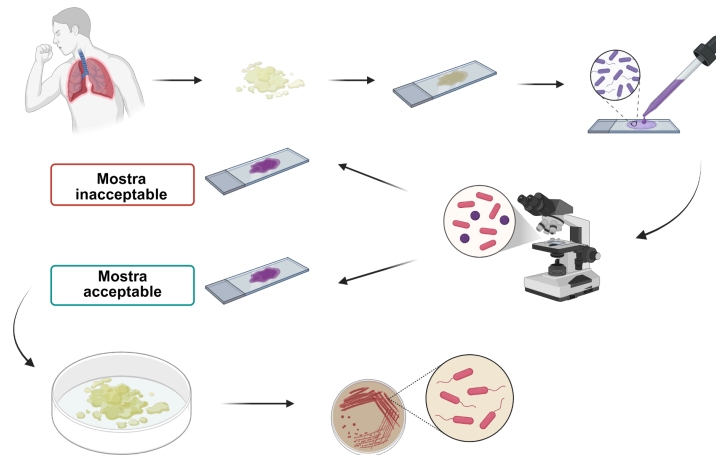


Figura 1: Flux de treball pel diagnòstic d'infeccions del tracte respiratori inferior. Creat a <https://BioRender.com>

3. Qualitat de la mostra i valor diagnòstic

El cultiu és la tècnica de referència per el diagnòstic d'ITRIs, però presenta limitacions importants, ja que la seva sensibilitat és moderada (40%-60%) (Cacho et al., s.d.). Una de les causes principals és que la mostra d'esput està exposada a un risc alt de contaminació amb la flora orofaríngea durant l'expectoració. Aquesta flora, pròpia de la boca i la gola, pot contenir microorganismes i biomarcadors que no són propis de les vies respiratòries inferiors. Quan aquestes espècies proliferen al cultiu,

poden ocultar els patògens responsables de la infecció, interferint en el diagnòstic microbiològic i donant lloc a falsos negatius. Com a conseqüència, es poden produir diagnòstics erronis, retards terapèutics i manca de la detecció d'infeccions greus. Un diagnòstic tardà pot comportar complicacions severes, tals com falles respiratòries, xoc sèptic o disfunció multiorgànica, incrementant el risc de mortalitat. Així com afavorir el desenvolupament de resistències antimicrobianes per una prescripció errònia d'antibiòtics (Bouzada et al., 2025). A més, el processament d'espècimens de mala qualitat pot comportar un malbaratament de recursos del laboratori com ara personal tècnic i reactius (Budayanti et al., 2019).

Cal destacar que, a nombrosos articles recents, es remarca la importància d'assegurar una qualitat adequada de les mostres, ja que s'ha detectat diferències significatives entre els resultats obtinguts a partir de mostres acceptables i aquelles considerades inacceptables. Les mostres de bona qualitat permeten aïllar un major nombre d'organismes potencialment patògens, al contrari que les mostres considerades inacceptables (Budayanti et al., 2019; Johny i Thomas, 2024).

En aquest context, la tinció de Gram s'ha consolidat com la tècnica de referència per avaluar la qualitat de l'esput. Es tracta d'un mètode ràpid, senzill i econòmic que permet observar els tipus de cèl·lules de la mostra a la vegada que facilita una primera aproximació a la morfologia dels bacteris presents. La presència abundant de leucòcits polimorfonuclears indica una resposta inflamatòria i una possible infecció activa, mentre que un excés de cèl·lules epitelials escamoses suggereix contaminació amb saliva i secrecions orofaríngees. Així, la relació entre aquests dos tipus cel·lulars constitueix la base dels criteris estandarditzats que determinen si una mostra presenta una qualitat acceptable o no. Entre ells, els criteris de Murray i Washington (1975) són dels més utilitzats i consideren vàlida una mostra que presenti més de 25 leucòcits per camp (100x) i menys de 10 cèl·lules epitelials escamoses per camp (100x) (Cacho et al., s.d.).

A més d'avaluar la qualitat, l'examen microscòpic de les mostres d'esput mitjançant tinció de Gram pot oferir pistes preliminars sobre l'agent etiològic de la infecció. Quan en una mostra acceptable es detecta un morfotip bacterià clarament

predominant, es pot suggerir l'organisme responsable de la infecció. Aquest fet reforça la utilitat d'aquesta tècnica com a pas inicial en el procés diagnòstic.

3.1. Tinció de Gram

La tinció de Gram constitueix el primer pas analític després de la recollida de l'esput i abans del cultiu. Tot i ser una tècnica senzilla, és essencial, ja que no només funciona com a eina de classificació bacteriana, sinó que també actua com a filtre de qualitat que determina quines mostres poden aportar un resultat clínicament fiable. Aquest mètode, desenvolupat per Hans Christian Gram fa més de 100 anys, comporta un procediment molt senzill.

El procediment comença amb l'aplicació del cristall violeta com a tinció primària, unint-se a la paret cel·lular dels bacteris de la mostra. Posteriorment, s'aplica una solució de iode que actua com a mordent, fixant el colorant a les cèl·lules. El següent pas de decoloració, amb alcohol o acetona, és determinant ja que elimina el colorant dels bacteris Gram negatius, però no dels positius, a causa de les diferències en la composició de la seva paret cel·lular. Finalment, els bacteris Gram negatius decolorats són tenyits amb una tinció de contrast, com la safranina, i apareixen vermells quan són observats al microscopi, mentre que els bacteris Gram positius mantenen un color violeta intens (Gary et al, 2006).

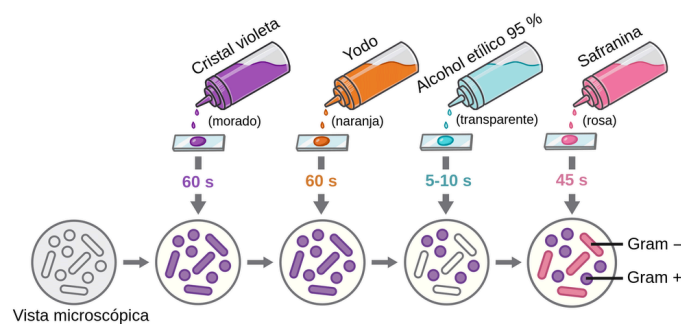


Figura 2: Procediment de la tinció de Gram. Steps to perform gram-staining. Labster. Disponible a: <https://theory.labster.com/steps-gramstain/> [accedit 30 Agost 2025].

3.2. Criteris d'avaluació

Com s'ha esmentat anteriorment, la presència de contaminació a la mostra pot comprometre la fiabilitat del diagnòstic microbiològic. Per tal de minimitzar aquest risc, és imperatiu establir un criteri estandarditzat per descartar espècimens inadequats a tots els laboratoris.

El criteri emprat en el present treball es centra en el sistema proposat per Murray i Washington (MW), que categoritza les mostres en sis grups amb diferents proporcions de cèl·lules epitelials i leucòcits, com es mostra a la *Taula 1*. Els grups 1 a 3 es consideren inacceptables, mentre que els grups 4 i 5 es poden acceptar com a mostres de bona qualitat. Per un altre costat, el grup 6, que presenta menys de 25 cèl·lules epitelials i leucòcits, es considera una mostra de poca qualitat.

Els grups amb un alt nombre de cèl·lules epitelials, com els grups 1 a 3 d'aquest sistema, indiquen contaminació per secrecions orofaríngees i invalida les mostres. L'objectiu final d'aplicar els criteris de MW és augmentar la probabilitat d'aïllar bacteris patògens de les mostres d'esput, millorant així la precisió del diagnòstic i l'eficàcia del tractament posterior.

Taula 1. Criteri clàssic per avaluar la qualitat d'un espècimen d'esput. Adaptat de: *Bacteriologic diagnosis of respiratory tract infection*. Figura de ResearchGate. Disponible a: https://www.researchgate.net/figure/Classical-criteria-for-evaluation-of-cellular-quality-of-a-sputum-specimen_tbl1_11494972 [accedit 30 Agost 2025].

	Cèl·lules epitelials per camp baix d'augment	Leucòcits per camp baix d'augment
Grup 1	> 25	< 10
Grup 2	> 25	10-25
Grup 3	> 25	> 25
Grup 4	10-25	> 25
Grup 5	< 10	> 25
Grup 6	< 25	< 25

4. Limitacions dels mètodes convencionals

Un dels principals reptes dels mètodes tradicionals com la tinció de Gram és la subjectivitat en la lectura de les mostres, ja que els resultats poden variar considerablement entre observadors depenent de la seva experiència i entrenament. En aquest cas, un observador casual pot no detectar els mateixos detalls que un professional experimentat, el que pot conduir a errors en la interpretació de la qualitat de l'esput (Gary et al., 2006). Un altre repte és la necessitat de reduir el temps dedicat a l'anàlisi, especialment en situacions en què el tractament antibiòtic és urgent, ja que la tinció Gram, tot i ser ràpida, no permet una resposta immediata i requereix personal qualificat i el temps suficient per obtenir resultats fiables.

5. Intel·ligència artificial

La intel·ligència artificial (IA) és una tecnologia cada vegada més present en múltiples àmbits, incloent-hi el camp de la microbiologia clínica. En l'àmbit clínic, la integració de la IA pot afavorir una pràctica més eficient i precisa, a més de facilitar l'extracció de conclusions que poden passar desapercebudes en l'anàlisi convencional. Aquest potencial suposa una oportunitat per millorar l'atenció als pacients (Smith et al., 2020).

Una de les aplicacions més destacades és l'anàlisi d'imatges basada en IA, el que permet introduir algorismes de *machine learning* (ML) per identificar patrons rellevants en mostres microbiològiques. Aquests algorismes, per si sols, no són capaços de reconèixer quines característiques d'una imatge són rellevants per la classificació o el diagnòstic. Per aquest motiu, sovint requereixen una supervisió humana, especialment durant la fase d'entrenament. Aquest procés, anomenat aprenentatge supervisat, implica la guia d'un expert que categoritza i etiqueta prèviament un conjunt d'imatges, conegut com a *training set*. El model, a partir d'aquestes etiquetes, genera un algorisme que intenta predir la categoria de noves imatges i compara els seus resultats amb les categories assignades originals per mesurar-ne la precisió (Smith et al., 2020).

Durant l'aprenentatge, els models poden desenvolupar associacions incorrectes si es fixen en trets no rellevants de les imatges. Aquest fenomen es coneix com a sobreajustament o *overfitting* i es dona quan l'algorisme s'entrena amb un conjunt de dades insuficient o poc divers. El sobreajustament sovint es detecta mitjançant l'anàlisi del conjunt de prova o *test set*, que està format per imatges que l'algorisme no ha analitzat durant l'entrenament i que serveixen per avaluar el seu rendiment en situacions reals (Smith et al., 2020).

5.1. Deep learning

Una de les estratègies més efectives per evitar el problema del sobreajustament és utilitzar conjunts de dades grans i variades. Per aquest propòsit, l'aprenentatge profund o *deep learning* (DL) ofereix avantatges, ja que pot analitzar milers o milions d'imatges per solucionar aquest problema. Aquest aprenentatge es basa en l'ús de xarxes neuronals, que són estructures amb diverses capes successives de processament inspirades en la connectivitat del còrtex visual humà (LeCun et al., 2015).

Els mètodes de *deep learning* constitueixen una forma avançada d'aprenentatge de representacions. Es basa en una arquitectura jeràrquica formada per múltiples mòduls de representació simples, però no lineals, que transformen les dades d'entrada de manera progressiva en representacions cada vegada més detallades. Amb el suficient nombre de transformacions els models poden aprendre funcions altament complexes (LeCun et al., 2015).

5.2. Xarxes neuronals convolucionals

Les xarxes neuronals convolucionals (CNN) són un tipus especialitzat de DL dissenyat per processar dades que venen en forma de múltiples matrius, com les imatges. Aquestes s'han aplicat exitosament a la detecció, segmentació i reconeixement d'objectes i regions d'imatges (LeCun et al., 2015).

L'arquitectura típica d'una CNN s'organitza en una sèrie de capes consecutives. Les primeres estan compostes habitualment per dos tipus de capes: les capes

convolucional i les d'agregació o *pooling*. Les capes convolucional són les encarregades d'extreure característiques locals, com vores o textures, mitjançant l'aplicació de filtres i les capes d'agregació són les que redueixen la dimensió dels mapes de característiques. Aquesta disminució no redueix la càrrega computacional i aporta robustesa davant de petites variacions, com canvis en la posició, la il·luminació o el fons dels elements visuals (LeCun et al., 2015). Una arquitectura molt utilitzada consisteix en una o dues capes convolucional, seguides d'una capa d'agregació.

Aquestes capes convolucional i d'agregació són responsables d'extreure les característiques bàsiques d'una imatge, generant un vector format per matrius de característiques. Per classificar la imatge, aquest vector s'aplana i s'introdueix en una xarxa neuronal totalment connectada (FCN), que calcula la probabilitat que la imatge pertanyi a diferents categories. Finalment, aplicant la funció softmax a l'última capa, es converteixen aquestes dades en una sola probabilitat clara, assignant la imatge a la categoria amb probabilitat més alta (Jeczminek i Kowalski, 2021).

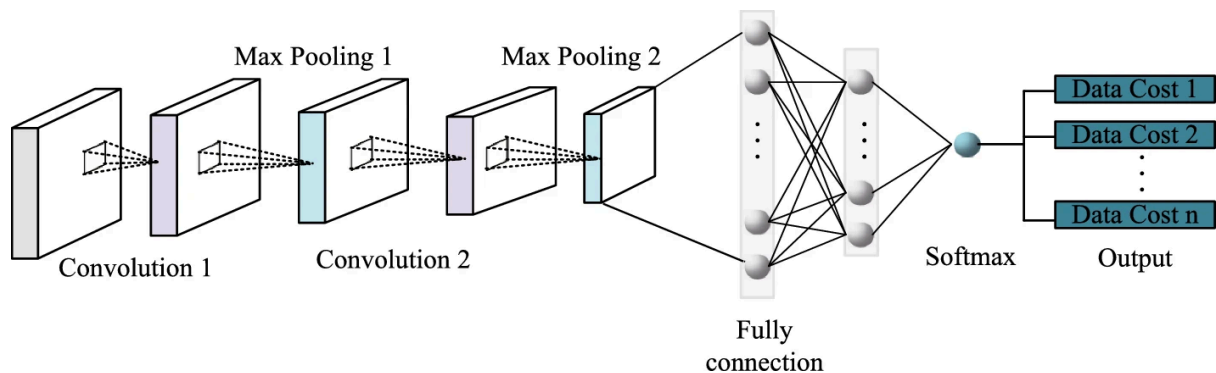


Figura 3: Estructura d'una CNN. Zhao, X., Wang, L., Zhang, Y. et al. A review of convolutional neural networks in computer vision. *Artif Intell Rev* 57, 99 (2024). Disponible a: <https://doi.org/10.1007/s10462-024-10721-6> [accedit 30 Agost 2025].

5.3. Tasques de visió per computador i aplicació clínica

Com s'ha mencionat anteriorment, les xarxes d'aprenentatge profund han contribuït de manera significativa a múltiples aplicacions com reconeixement d'imatges i classificació, detecció d'objectes i segmentació d'imatges.

La classificació d'imatges permet identificar a quina categoria pertany una imatge completa. En aquest procés, l'algorisme analitza el contingut global de la imatge i assigna una única etiqueta que descriu la classe d'aquesta. Un exemple són els sistemes de CAPTCHA, on es demana distingir entre imatges amb senyal d'aturada i sense (Murel i Kavlakoglu, 2024).

La detecció d'objectes no només determina si un tipus d'objecte és present a una imatge, sinó que també identifica on es troba. Mentre que la classificació etiqueta tota la imatge, la detecció dibuixa caixes limitadores que indiquen la posició i la classe de cada objecte detectat. Per exemple, un sistema de detecció pot reconèixer totes les senyals de trànsit, cotxes i vianants presents indicant la posició exacta (Murel i Kavlakoglu, 2024).

La segmentació d'imatges ofereix encara més detall en la identificació d'objectes. Aquesta tècnica, en lloc d'utilitzar caixes limitadores, divideix una imatge en múltiples segments o regions, cadascuna formada per conjunts de píxels que comparteixen característiques comunes, com el color, la textura o la intensitat. Això permet localitzar contorns, línies i altres estructures rellevants de la imatge, resultant especialment útil en aplicacions com la detecció d'objectes o l'anàlisi biomèdica (Murel i Kavlakoglu, 2024).

Un exemple concret de l'ús de CNN en microbiologia clínica és l'ús d'aquestes per a la classificació de morfologies bacterianes en tincions de Gram de cultius de sang positius. En aquest cas, a partir de la imatge digitalitzada del portaobjectes, es retallen múltiples regions petites, anomenades *crops*, que contenen morfologies bacterianes específiques. Els fragments representen subimatges localitzades del portaobjectes i s'utilitzen com a unitats d'entrenament per al model, que aprèn a identificar patrons visuals i classificar-los automàticament. Aquest mètode va aconseguir una precisió superior al 95% en classificació de *crops* a l'estudi de Smith et al., 2020.

5.4. Limitacions i consideracions per la implementació de la IA a l'àmbit mèdic

Malgrat els avenços significatius assolits per la intel·ligència artificial en diversos àmbits mèdics i diagnòstics, encara hi ha diversos reptes per aconseguir una implementació efectiva.

En primer lloc, gran part de la investigació sobre l'impacte de la IA en l'atenció mèdica es du a terme en entorns experimentals i controlats, fora de la pràctica clínica habitual, el que dificulta la validació de la seva eficàcia. A més, la implementació de la IA també requereix que s'integri perfectament als fluxos de treball dels professionals existents, ja que les innovacions han de millorar l'eficiència, en lloc d'alentir els processos assistencials. Cal tenir en compte que la implementació de tecnologies d'intel·ligència artificial requereix temps i recursos per capacitar als professionals implicats (Aung et al., 2021).

D'altra banda, donada l'absència d'un consens establert sobre la regulació de la intel·ligència artificial, actualment no és viable dissenyar sistemes algorísmics que es puguin generalitzar de manera uniforme a tots els entorns sanitaris. Per aquesta raó, Aung et al. proposen que pot ser més prudent centrar-se en el desenvolupament de sistemes adaptats a les característiques i necessitats específiques de les institucions per les quals han estat dissenyats.

L'adopció de la IA en l'àmbit sanitari serà un procés progressiu, lent i acuradament supervisat que requerirà una regulació estricta i un seguiment constant de la seva utilitat i eficàcia. No obstant això, amb la supervisió adequada la IA pot esdevenir una eina complementària molt valuosa per millorar el benestar dels pacients.

HIPÒTESIS DE TREBALL I OBJECTIUS

L'avaluació de la qualitat de les mostres d'esput mitjançant la tinció de Gram és una tècnica àmpliament utilitzada en microbiologia clínica, ja que permet observar de manera ràpida i econòmica els tipus de cèl·lules presents i obtenir una primera aproximació a la morfologia dels bacteris. No obstant això, per un costat, aquesta tècnica és molt subjectiva, ja que depèn de l'experiència i l'habilitat de l'observador, el que pot provocar variabilitat i errors en la classificació. Per l'altre costat, encara que és un procediment relativament ràpid, la lectura i interpretació de les mostres requereix temps i personal qualificat, el que pot retardar l'inici del tractament.

Davant d'aquesta problemàtica, hem plantejat el desenvolupament d'un sistema automatitzat que sigui capaç de processar imatges microscòpiques d'esput tenyides i classificar-les segons la seva qualitat. D'aquesta manera, es podria reduir la subjectivitat de la lectura, estandarditzar la interpretació i escurçar el temps de resposta en entorns hospitalaris, on moltes vegades la rapidesa en la presa de decisions és crucial.

Hipòtesis

És tècnicament viable desenvolupar un sistema basat en xarxes neuronals convolucionals capaç de detectar i comptar automàticament cèl·lules epitelials i leucòcits en imatges microscòpiques d'esput amb l'objectiu de classificar-les segons la seva qualitat de manera objectiva i ràpida.

Objectiu principal

L'objectiu principal del present treball és desenvolupar i avaluar, com a prova concepte, la viabilitat tècnica d'un sistema automatitzat capaç de determinar la qualitat de mostres d'esput segons els criteris de Murray i Washington mitjançant la detecció i el comptatge de cèl·lules epitelials i leucòcits utilitzant una xarxa neuronal entrenada en un entorn local. Aquest sistema, complementat amb una interfície gràfica per facilitar el seu ús en entorns clínics, busca reduir la subjectivitat i estandarditzar aquest procés.

METODOLOGIA

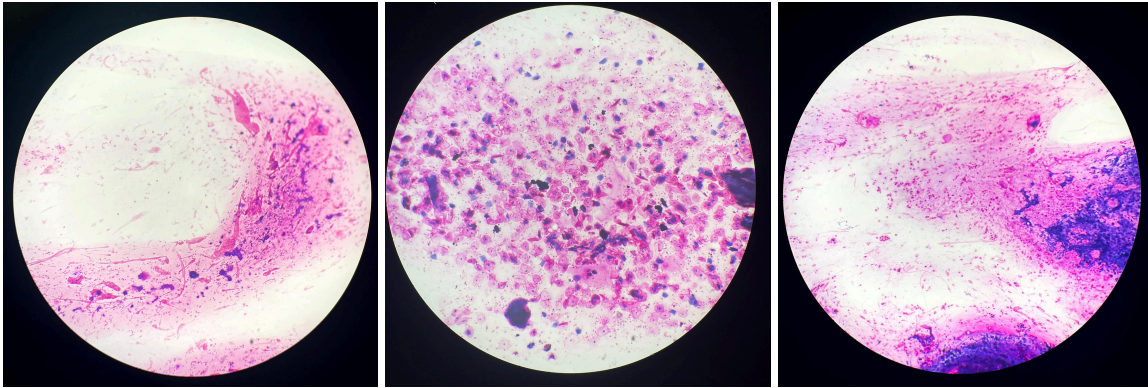
Abans d'iniciar l'adquisició d'imatges i el processament computacional, vaig dur a terme una fase de formació amb l'objectiu de familiaritzar-me amb les mostres d'esput observades sota el microscopi òptic. Aquesta etapa em va permetre reconèixer les diferents categories de qualitat segons la presència i proporció de cèl·lules epitelials i leucòcits polimorfonuclears.

1. Adquisició de les imatges

Les mostres d'esput utilitzades en aquest estudi van ser proporcionades pel laboratori d'infeccions respiratòries del Servei de Microbiologia del Centre de Diagnòstic Biomèdic de l'Hospital Clínic de Barcelona. Aquestes mostres van ser tenyides mitjançant el mètode de tinció de Gram, seguint un protocol estandarditzat que assegurava un contrast cromàtic òptim i una il·luminació homogènia al llarg de totes les preparacions.

L'adquisició d'imatges de les mostres es va realitzar mitjançant un microscopi òptic equipat amb una càmera digital d'un dispositiu mòbil iPhone 13. Per tal de garantir una alineació adequada i estabilitat durant la captura d'imatges, es va utilitzar un adaptador per connectar el dispositiu mòbil a l'ocular del microscopi, el que va permetre registrar imatges d'alta qualitat de les mostres tenyides. Les imatges es van obtenir amb l'objectiu de 10x, la configuració habitual en l'avaluació microbiològica de la qualitat de preparacions, atès que proporciona un camp visual adequat per identificar la presència i proporció de cèl·lules epitelials i leucòcits.

Es van adquirir imatges de 280 mostres, capturant entre quatre i cinc camps per mostra. Posteriorment, es van seleccionar aquelles fotografies que presentaven millor qualitat d'enfocament i que eren més representatives, obtenint finalment un conjunt de dades final de 402 imatges. Les microfotografies resultants, de resolució 640x640 píxels, es van organitzar en tres categories segons la qualitat de la preparació: BQ (bona qualitat), MQ (mala qualitat) i PQ (poca qualitat).



Figures 4,5 i 6: Algunes de les imatges de les mostres d'esput recollides.

Tot i que el criteri de Murray i Washington (MW) categoritza les mostres en sis grups segons la proporció de leucòcits i cèl·lules epitelials, per aquest projecte s'ha optat per realitzar una reagrupació en tres classes funcionals (BQ, MQ i PQ). Aquesta decisió es basa en la pràctica habitual observada al laboratori on s'ha desenvolupat l'estudi, on els grups 4 i 5 es consideren de bona qualitat (BQ), els grups 1, 2 i 3 s'agrupen com a mala qualitat (MQ) i el grup 6 s'interpreta com a mostra de poca qualitat (PQ). Amb aquesta agrupació s'ha buscat simplificar la tasca de classificació mantenint la coherència amb els criteris clínics utilitzats per descartar mostres no adequades i facilitar la implementació d'un model de *deep learning* amb classes clarament diferenciades.

2. Anotació del conjunt de dades

Abans d'iniciar l'entrenament, totes les imatges es van anotar utilitzant la plataforma Roboflow Classical. Roboflow és una eina en línia àmpliament utilitzada en visió per computador que permet organitzar, anotar i preparar conjunts de dades d'imatges per a l'entrenament de models d'aprenentatge automàtic.

En aquest projecte es va crear un conjunt de dades amb dues etiquetes principals, MQ (mala qualitat) i BQ (bona qualitat). La tasca d'anotació es va dur a terme mitjançant l'eina de segmentació d'instàncies, una tècnica que va permetre dibuixar manualment les màscares que delimitaven amb precisió les regions corresponents de cada classe.

La segmentació per instàncies permet identificar cada objecte d'interès de manera individual i delimitar els contorns píxel-a-píxel. A diferència de la segmentació semàntica, que assigna la mateixa etiqueta a tots els objectes d'una classe sense distingir-los, la segmentació per instàncies tracta separatament cada mostra dins d'una mateixa categoria (Contributing Writer, 2025).

Per cada categoria, es va reservar un 60% de les imatges per l'entrenament, un 20% per la validació i un 20% per prova, mantenint la mateixa proporció de classes a cada partició. Les anotacions exportades des de Roboflow per a les classes MQ i BQ es van integrar directament a les carpetes *train/labels*, *valid/labels* i *test/labels*, acompanyades de les seves corresponents imatges a *train/images*, *valid/images* i *test/images*.

Finalment, el conjunt anotat es va exportar en format YOLOv8-seg, generant tant els arxius d'imatge com les seves etiquetes en .txt.

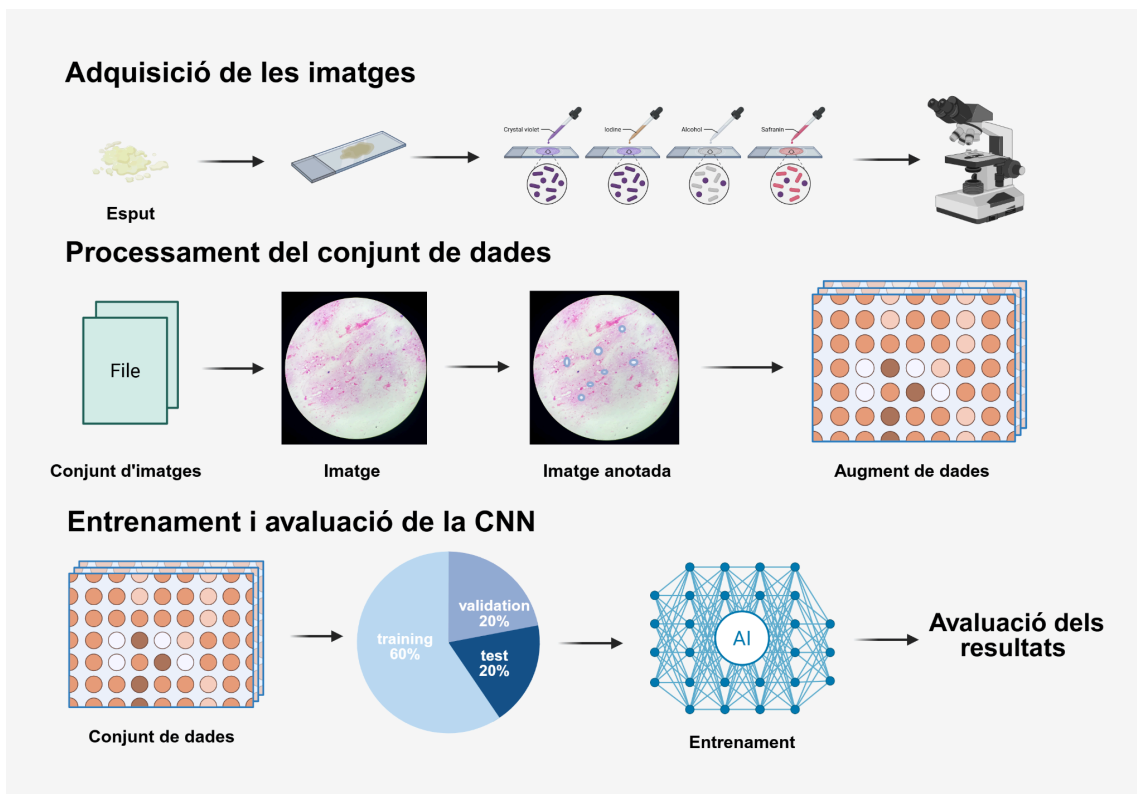


Figura 7: Esquema del flux de treball. Creat a <https://BioRender.com>.

3. Arquitectura de la CNN

En el moment de començar aquest projecte no comptava amb coneixements previs sobre el funcionament intern o el disseny de xarxes neuronals convolucionals. És per això que, degut a la complexitat que suposa desenvolupar una arquitectura pròpia, es va considerar més adequat partir d'un model ja establert i reconegut per la seva robustesa. Amb el suport i orientació d'un especialista en el camp es va seleccionar i adaptar l'arquitectura YOLOv8-seg, ajustant-la a les necessitats específiques del treball.

YOLOv8 representa una de les darreres evolucions de la sèrie You Only Look Once, reconeguda per combinar una alta velocitat de processament amb una precisió competitiva en la detecció d'objectes. A diferència d'altres models de detecció d'objectes, els models YOLO són de tipus One-Stage Detector, és a dir, que realitza la detecció i classificació en una sola etapa. YOLOv8, desenvolupat per l'equip d'Ultralytics, amplia les capacitats de les versions anteriors incorporant millores clau que reforcen el rendiment en escenaris de processament en temps real (Torres, 2025).

A la Figura 8 es mostra de manera esquemàtica l'arquitectura general dels models de detecció d'objectes, destacant les diferències entre els detectors *one-stage* i *two-stage*.

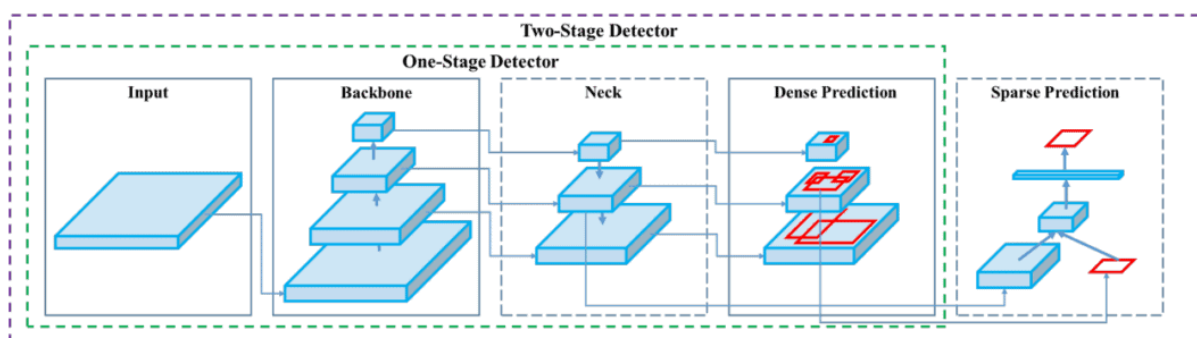


Figura 8: Mecanisme essencial d'un model de detecció d'objectes. Explore YOLOv8: Latest in Object Detection Tech. (s. d.). viso.ai. Disponible a: <https://viso.ai/deep-learning/yolov8-guide/> [accedit 30 Agost 2025].

Per a la segmentació d'instàncies i la classificació de la qualitat de les imatges, el model emprat va ser una variant lleugera de YOLOv8-seg (You Only Look One, versió 8). Aquest model és capaç de detectar i segmentar objectes en una sola passada (*single-stage inference*), amb una taxa superior als 30 fotogrames per segon (fps). Aquest rendiment el posiciona com una de les arquitectures més avançades en l'àmbit de la detecció d'objectes, destacant per la seva velocitat d'inferència en temps real i una precisió competitiva en entorns amb recursos computacionals limitats.

Una innovació fonamental de YOLOv8 és el seu sistema de detecció sense àncora, que elimina la necessitat d'utilitzar caixes prèviament definides per localitzar els objectes. En versions anteriors de YOLO aquestes caixes anomenades àncora eren imprescindibles i requerien un ajustament basat en el coneixement previ sobre les mides i proporcions dels objectes a detectar. Aquesta detecció sense àncora permet millorar el rendiment del model, ja que aporta una millor capacitat de generalització, especialment quan es treballa amb conjunts de dades on els objectes tenen configuracions inusuals o irregulars, com és el cas de les cèl·lules de les mostres segmentades en aquest projecte (Torres, 2025).

L'arquitectura de YOLOv8 integra els avenços recents en aprenentatge profund mitjançant la combinació de CNN amb estructures de piràmides de característiques. Això permet capturar simultàniament el context global de la imatge i els detalls més fins, potenciant la capacitat del model per afrontar tasques de segmentació d'instàncies d'elevada complexitat. L'extracció de característiques en múltiples escales facilita la identificació precisa dels límits dels objectes i la generació de màscares de segmentació d'alta qualitat (Torres, 2025).

YOLOv8-seg presenta una arquitectura modular (*Figura 9*) organitzada en tres components principals: *backbone*, *neck* i cap desacoblat (*head*).

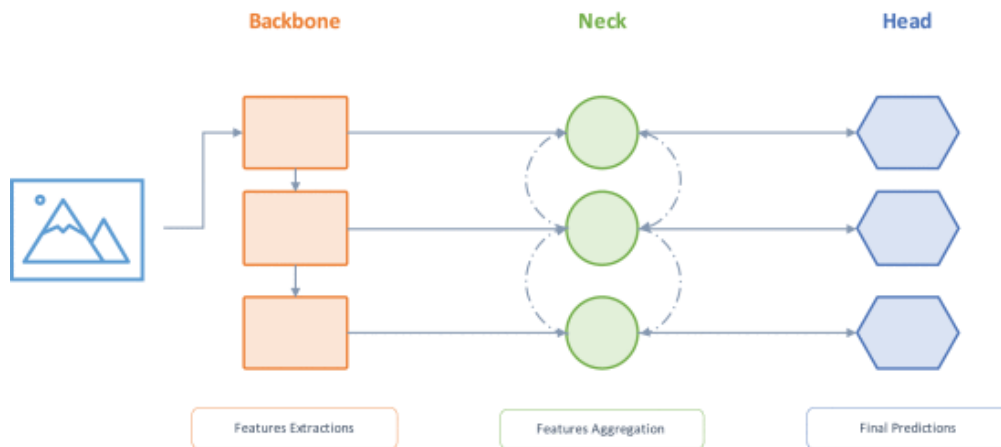


Figura 9: Arquitectura general dels models YOLO, que consisteixen en un backbone, neck i head. M. G. Ragab et al., "A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023)," in *IEEE Access*, vol. 12, pp. 57815-57836, 2024. Disponible a: <https://ieeexplore.ieee.org/abstract/document/10494845> [accedit 30 Agost 2025].

El *backbone* del model s'encarrega de l'extracció de característiques a partir de la imatge d'entrada mitjançant capes convolucionals profundes. Per això, s'ha utilitzat CSPDarknet, una arquitectura derivada de Darknet que incorpora connexions parcials entre etapes (Cross-Stage Partial Connections, CSP). Aquestes connexions afavoreixen una millor propagació de la informació durant l'entrenament i milloren la reutilització d'informació entre capes (Torres, 2025).

La secció intermitja o *neck* utilitza una *Path Aggregation Network* (PAN-FPN), que combina informació provinent de diferents nivells de profunditat del *backbone*. Aquesta arquitectura està dissenyada per facilitar el flux d'informació entre diferents resolucions espacials, fet que permet al model capturar de manera efectiva característiques a múltiples escales i millorar la detecció d'objectes a diverses mides (Torres, 2025).

El cap final (*head*) es troba desacoblat en dues branques independents. Una branca de classificació, basada en convolucions 1x1, assigna la categoria de qualitat (BQ, MQ o PQ) a cada regió detectada. I una branca de segmentació, que utilitza convolucions transposades per reconstruir la forma original dels objectes i blocs d'atenció espacial per focalitzar el processament en les àrees més rellevants, millorant la precisió del contorn de les màscares segmentades (Torres, 2025).

Aquest cap desacoblat incorpora millores respecte a versions anteriors de YOLO, com una funció de pèrdua basada en *Intersection over Union* (IoU), que millora la gestió d'objectes superposats i la precisió de les caixes delimitadores (Torres, 2025).

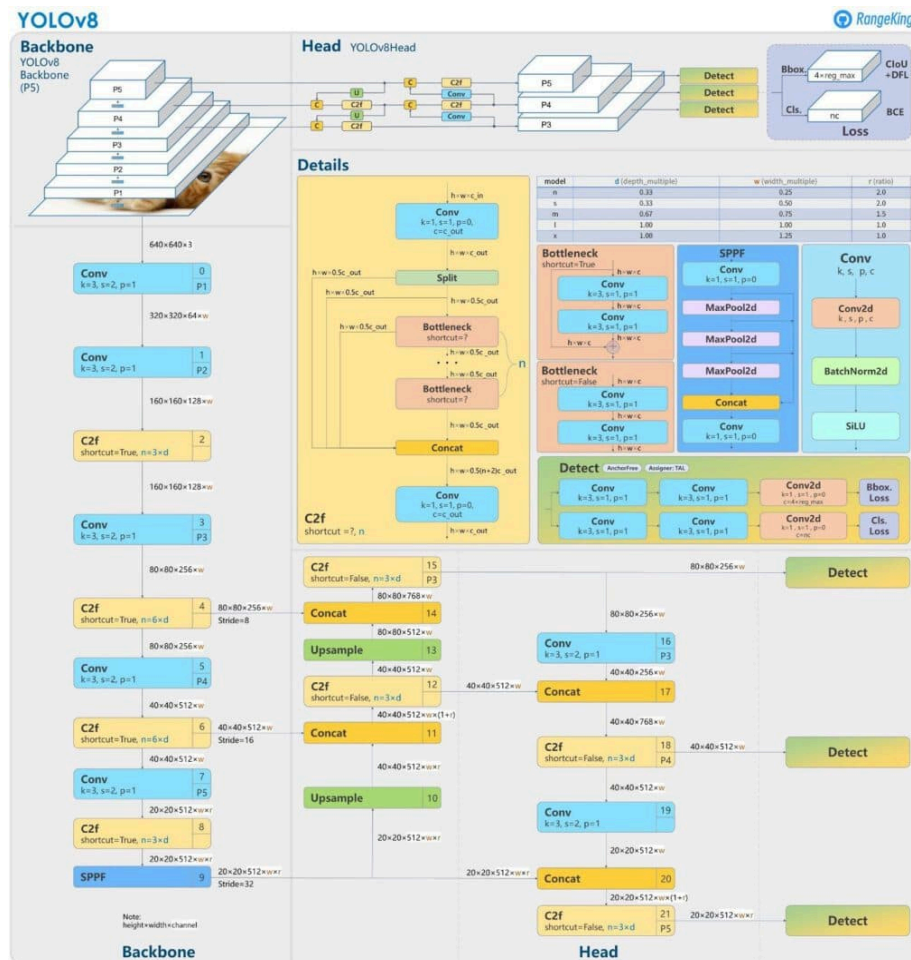


Figura 10: Arquitectura del model YOLOv8. Explore YOLOv8: Latest in Object Detection Tech. (s. d.). viso.ai. Disponible a: <https://viso.ai/deep-learning/yolov8-guide/> [accedit 30 Agost 2025].

4. Estratègia d'entrenament de la CNN

En els models de segmentació l'objectiu principal és assignar una categoria específica a cada píxel de la imatge. Per avaluar com de bé ho fa el model s'utilitza la funció de pèrdua, que quantifica la diferència entre les etiquetes predites pel model i les etiquetes reals (veritat fonamental) en cada píxel. Aquesta mesura serveix per quantificar el rendiment i a més juga un paper fonamental en guiar

l'aprenentatge del model, ajudant-lo a millorar les seves prediccions de manera iterativa.

L'elecció d'una funció de pèrdua adequada depèn de les característiques específiques del problema, com el desequilibri entre classes o la necessitat d'una alta precisió en la segmentació.

L'entrenament del nostre model es va realitzar durant 50 èpoques en una estació de treball equipada amb GPU (unitat de processament gràfic), emprant un conjunt de dades equilibrat per les classes MQ i BQ. En aquest cas es va utilitzar una funció de pèrdua composta, dissenyada per optimitzar simultàniament tres aspectes:

- Regularitzador de IoU, que mesura la similitud entre la segmentació predita pel model i la segmentació real. Es calcula com 1 menys l'índex de Jaccard, conegut també com a *Intersection over Union* (IoU), que quantifica el grau de solapament entre les àrees predites (P) i les reals (GT) (Figura 11). Aquest afavoreix una millor alineació espacial de les caixes de detecció respecte a les anotacions reals (Pratiher, 2024).

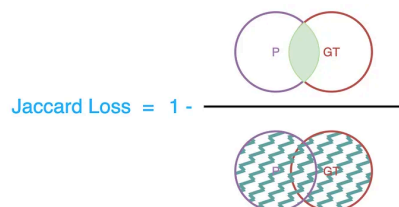


Figura 11: Funció de la IoU. Pratiher, D. (2024, 12 maig). *Understanding loss functions for deep learning segmentation models*. Medium. Disponible a:

<https://medium.com/@devanshipratiher/understanding-loss-functions-for-deep-learning-segmentation-models-30187836b30a> [accedit 30 Agost 2025].

- Pèrdua Dice (puntuació F1) és similar a l'anterior, però resulta especialment útil en conjunts de dades amb desequilibri de classes, és a dir, quan hi ha regions o categories que apareixen molt més que d'altres. Aquesta funció (Figura 12) tracta de manera equilibrada els falsos positius i els falsos negatius, sent sensible a la mida de la intersecció entre les segmentacions predites i les reals. En essència, és equivalent a utilitzar la puntuació F1 com

a funció de pèrdua, el que la fa molt eficaç per optimitzar models en escenaris amb classes descompensades (Pratiher, 2024).

$$\text{Dice Loss} = \frac{2 \times \text{P} \cap \text{GT}}{\text{P} \cup \text{GT}}$$

Figura 12: Funció per la pèrdua Dice. Pratiher, D. (2024, 12 maig). *Understanding loss functions for deep learning segmentation models. Medium. Disponible a: <https://medium.com/@devanshipratiher/understanding-loss-functions-for-deep-learning-segmentation-models-30187836b30a> [accedit 30 Agost 2025].*

- L'entropia creuada ponderada és una funció de pèrdua que mesura les diferències píxel a píxel entre les probabilitats predites del model i les etiquetes binàries reals (GT) en una imatge. Aquesta s'utilitza per millorar la classificació entre categories de qualitat (Pratiher, 2024).

Per a l'optimització es va fer servir l'algorisme AdamW amb un paràmetre de *weight decay* de 0,0005 per evitar un sobreajustament. La taxa d'aprenentatge es va ajustar dinàmicament mitjançant un *scheduler One-Cycle*, que la va augmentar des de 10^{-4} fins a 10^{-2} durant la fase d'escalfament i posteriorment la va reduir de forma progressiva, mantenint un momentum variable entre 0,85 i 0,95.

4.1. Augment de dades

Degut a la disponibilitat limitada de mostres i amb l'objectiu d'incrementar la diversitat del *dataset* i de millorar la capacitat de generalització del model, es van aplicar tècniques d'augmentació de dades en temps real. Aquestes tècniques permeten generar noves variants de les imatges originals ampliant la base de dades sense necessitat de recollir més mostres i reduint el risc d'*overfitting* o sobreajustament. En concret, es van utilitzar:

- Mosaics de quatre imatges, que combinen quatre imatges originals en una sola i després s'apliquen diferents transformacions al mosaic. Això, a part de

diversificar les dades d'entrada, millora la capacitat del model de detectar i reconèixer objectes que poden aparèixer parcialment a les imatges originals (Torres, 2025).

- Cutmix que, en lloc de reemplaçar parcialment regions d'una imatge amb pegats negres o grisos, les reemplaça amb fragments d'una altra imatge per aportar variabilitat (Torres, 2025).
- Variacions aleatòries d'escala i brillantor, una tècnica d'augment de dades que consisteix en aplicar modificacions aleatòries a la mida i les condicions lumíniques de les imatges d'entrenament. La variació d'escala simula objectes a diferents distàncies de la càmera, permetent que el model aprengui a detectar objectes de mides molt diverses. I, per un altre costat, els ajustos de brillantor reproduïxen situacions d'il·luminació variables afavorint que el model mantingui un bon rendiment en entorns visuals canviants (Torres, 2025).

Aquest procés es va dur a terme en cada lot (batch) de grandària 16, amb la finalitat d'equilibrar les classes i optimitzar el rendiment global del model, incrementant la capacitat de generalització del model.

5. Interfície gràfica

Per facilitar l'ús clínic i investigador del sistema, es va desenvolupar una interfície gràfica d'usuari (GUI) mitjançant el mòdul Tkinter de Python. Aquesta eina permet a l'usuari carregar imatges individuals o directoris complets, executar la segmentació automàtica i visualitzar en temps real les deteccions obtingudes mitjançant marcadors centrats en cada instància. A més, la GUI mostra de manera destacada la classe de qualitat predominant detectada (MQ o BQ), oferint així una eina intuïtiva i reproduïble per inspeccionar la qualitat d'imatges microscòpiques en entorns mèdics o de recerca.

RESULTATS

La convergència del model MurrAI Quality es va supervisar mitjançant la mètrica IoU i es va validar periòdicament sobre el conjunt de validació utilitzant precisió, *recall* i *F1-score*. Un cop completat l'entrenament, el model es va avaluar sobre el conjunt de prova (20% del total), i es van analitzar els resultats.

Les mètriques de rendiment són eines essencials per avaluar amb precisió i objectivitat la capacitat d'un model de detecció d'objectes, ja que permeten mesurar la seva eficàcia per identificar i localitzar objectes i la manera en què gestiona errors habituals, com els falsos positius i falsos negatius. Aquestes mètriques no només són rellevants per a YOLOv8, sinó que també s'utilitzen en altres arquitectures de detecció d'objectes (Ultralytics, 2025).

Entre les més destacades trobem:

- Intersecció sobre Unió (IoU): mesura la superposició entre la caixa delimitadora predita i la real (veritat fonamental o *ground truth*), el que és clau per avaluar la precisió de la localització (Ultralytics, 2025). Aquesta mètrica es mesura en una escala de 0 a 1, on 1 representa el valor ideal (Shah, s.d.).

$$IoU = \frac{\text{Àrea de superposició}}{\text{Àrea total d'unió}} \quad (1)$$

- Precisió mitjana (AP): calcula l'àrea sota la corba de precisió - recuperació (*recall*), resumint en un únic valor l'equilibri entre la capacitat del model per encertar (precisió) i detectar totes les instàncies (recuperació) (Ultralytics, 2025).
- Precisió mitjana promig (mAP): calcula els valors AP promig en múltiples classes d'objectes, proporcionant una visió global del rendiment en deteccions multiclasse (Ultralytics, 2025). En aquest cas i en AP 1 equival a detectar totes les instàncies correctament (Shah, s.d.).

- Precisió i recuperació (*recall*): la precisió quantifica la proporció de prediccions veritables (PV) respecte a la suma de veritables i falsos positius (FP), mentre que la recuperació indica la proporció de positius veritables (PV) respecte a la suma de positius veritables i negatius falsos (FN) (Ultralytics, 2025).

$$Precisió = \frac{PV}{PV + FP} \quad (2) \qquad Recall = \frac{PV}{PV + FN} \quad (3)$$

- Puntuació F1 (*F1 score*): mitjana harmònica entre precisió i recuperació, que ofereix una visió equilibrada del rendiment tenint en compte tant falsos positius com falsos negatius (Ultralytics, 2025).

$$F1 \text{ score} = 2 \times \frac{Precisió \times Recall}{Precisió + Recall} \quad (4)$$

Per precisió, recuperació i puntuació F1 un valor de 1 representa una puntuació perfecta sense errors en les prediccions.

En el cas de YOLOv8 també són habituals altres mètriques específiques. La mAP50 avalua el rendiment amb un llindar d'IoU del 0,50, convertint-se en una mesura de la precisió en les deteccions considerades "fàcils". D'altra banda, la mAP50-95 calcula el promig de l'AP en un conjunt de llindars que van del 0,50 al 0,95, el que permet avaluar el model tant en situacions senzilles com en casos on es requereix una alta precisió (Ultralytics, 2025).

La precisió i la recuperació ajuden a entendre el balanç entre falsos positius i falsos negatius, mentre que la IoU i el mAP donen informació sobre la qualitat de la segmentació. Finalment, la puntuació F1 ofereix una avaluació completa del rendiment del model (Haifa Alanazi, 2025).

L'entrenament del model YOLOv8-seg va completar 50 èpoques amb una durada aproximada d'un minut, generant dos pesos principals (last.pt i best.pt) amb una mida de 6.2 MB cadascun. Durant la fase de validació, realitzada sobre un conjunt de 44 imatges que contenen 873 instàncies anotades, el rendiment global de

segmentació d'instàncies va ser baix, amb un mAP50 de 0,223 i un mAP50-95 de 0,110 (*Taula 2*).

L'anàlisi per classes va mostrar que la detecció de cèl·lules de mala qualitat (MQ) va assolir un mAP50 de 0,525, mentre que la detecció per a bona qualitat (BQ) i poca qualitat (PQ) va ser considerablement inferior, amb valors de mAP50 de 0,012 i 0,133 respectivament. La precisió global (P) i el *recall* global (R) van ser de 0,223 i 0,319 (*Taula 2*), amb un comportament clarament superior en la classe MQ (P = 0,224 i R = 0,830) en comparació amb BQ (P = 0,125 i R = 0,012).

Per simplificar, a la fase de classificació (*Taula 3*) es van reagrupar les classes MQ i PQ per avaluar un escenari binari enfront de BQ. En aquest cas, els resultats van millorar de manera notable, amb una precisió macro de 0,75, un *recall* macro de 0,95 i un *F1-score* macro de 0,81. La classe combinada MQ + PQ va assolir una precisió de 1,00 i un *recall* de 0,91 (F1 = 0,95), mentre que la classe BQ va obtenir una precisió de 0,50 i un *recall* de 1,00 (F1 = 0,67). L'exactitud global en classificació binària va ser del 92%, amb un *F1-score* ponderat de 0,93. Aquestes mètriques es van registrar i guardar automàticament a *classification_report_binario.csv*.

Visualment, les prediccions es van desar a *runs/train/murrai_v3113/predict*. En aquestes sortides, la segmentació de mostres MQ era àmpliament visible, mentre que les deteccions corresponents a BQ i PQ eren més escasses. La velocitat d'inferència, d'aproximadament 3 ms per imatge, confirma la viabilitat del model per a la seva integració en entorns amb requisits de processament en temps real.

Taula 2. Mètriques de rendiment del sistema.

Mètrica de segmentació	Valor
Precisió global (P)	0,223
<i>Recall</i> global (R)	0,319
mAP50	0,223
mAP50-95	0,110
Velocitat (inferència)	aprox 3 ms/imatge

Taula 3. Anàlisi de la classificació binària.

Classe	Precisió	Recall	F1-score	Suport
MQ + PQ	1,00	0,91	0,95	11
BQ	0,50	1,00	0,67	1
Global	0,92	0,92	0,93	12

L'avaluació detallada del rendiment posa de manifest diferències importants entre tasques de segmentació d'instàncies i la classificació binària de qualitat. En el cas de la segmentació, el valor de mAP50 obtingut (0,223) indica que només el 22,3% de les màscares generades presenten una superposició (IoU) igual o superior al 50% respecte a les anotacions de referència. Quan el llindar d'IoU s'incrementa fins al 95%, el mAP es redueix dràsticament fins a 0,110, fet que evidencia que, si bé el model és capaç de captar la ubicació general dels objectes, la precisió píxel-a-píxel encara és insuficient per a aplicacions que requereixen contorns ajustats amb elevada fidelitat. Per tant, aquest fet evidencia que actualment el sistema no pot substituir l'avaluació professional en tasques on es necessiti una segmentació molt precisa de les cèl·lules.

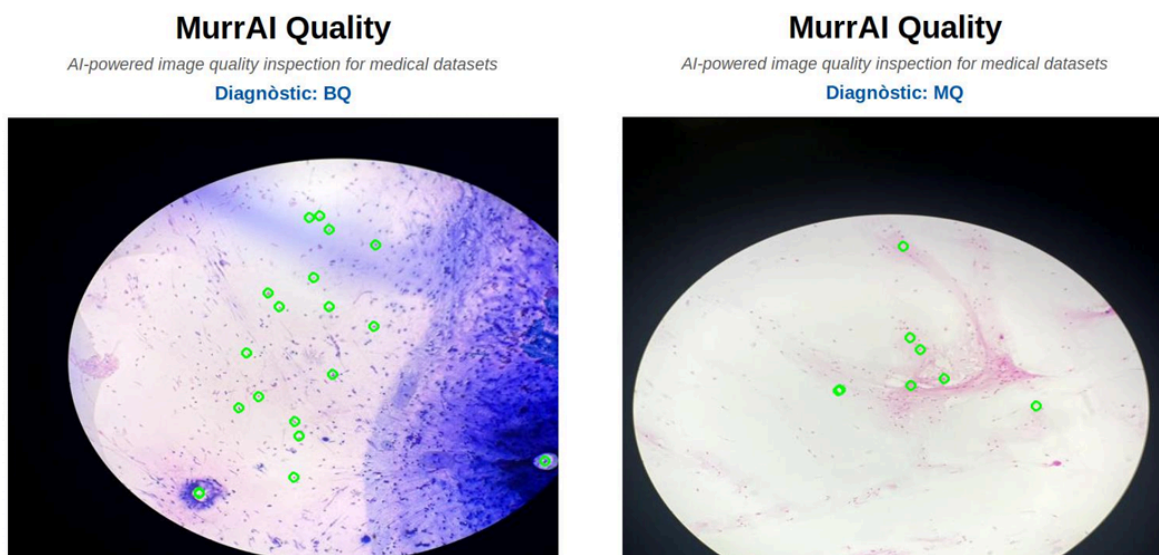
Aquest comportament també es veu reflectit en la precisió global (0,223) i el *recall* global (0,319), que confirmen que el detector presenta limitacions per identificar totes les regions d'interès amb un nivell de confiança alt.

Tot i això, la classe de mala qualitat aconsegueix un *recall* proper a 0,83, el que suggereix que la xarxa ha après de manera eficaç les característiques de les imatges de preparacions inacceptables i pot ser especialment útil com eina per descartar mostres que no compleixin amb els criteris mínims de qualitat. En canvi, el rendiment en la detecció de mostres de bona qualitat és significativament inferior (*recall* proper a 0,01), el que es deu principalment a la menor representativitat de la categoria BQ durant l'entrenament i a la similitud morfològica que presenten algunes instàncies de PQ i MQ. Clínicament, aquesta limitació implica que el sistema encara no està llest per utilitzar-se com a únic criteri per confirmar que una mostra és de bona qualitat i hauria d'anar acompanyat d'una validació manual de part d'un professional.

En un escenari simplificat de classificació binària, on es van agrupar les classes MQ i PQ enfront de BQ, el model va assolir una exactitud del 92% i un *F1-score* ponderat de 0,93, fet que reflecteix un bon equilibri entre precisió i *recall*. La classe combinada MQ + PQ va obtenir un *F1-score* de 0,95, sustentat per una precisió perfecta (1,00) i un *recall* elevat (0,91), el que garanteix que la majoria de mostres de qualitat inacceptable seran detectades correctament. En canvi, la classe BQ va mostrar un *recall* de 1,00 però amb una precisió de 0,50, indicant l'existència de falsos positius que es poden atribuir a un llindar de confiança poc restrictiu o a superposicions entre característiques de textura.

Des del punt de vista d'eficiència, la latència mitjana d'inferència es va situar al voltant dels 3 mil·lisegons (ms) per imatge en GPU, equivalent a més de 300 fps en condicions òptimes, cosa que permet considerar la solució viable per a entorns de processament en temps real. Aquest rendiment és especialment rellevant en contextos clínics, on la resposta immediata pot resultar crítica per a la presa de decisions i reforça el potencial d'aquest sistema com una eina de suport.

A les figures 13 i 14 es mostra la interfície gràfica d'usuari que s'ha desenvolupat, on es pot visualitzar la representació dels resultats de segmentació i classificació.



Figures 13 i 14: Interfície gràfica MurrAI Quality.

En conjunt, el sistema mostra un comportament especialment robust en la detecció de mostres de mala qualitat i en la classificació global de la qualitat, i presenta una capacitat operativa compatible amb fluxos de treball de gran demanda. Això implica que, clínicament, aquest sistema és molt eficaç com a eina de cribratge, ja que quan marca una mostra com a “MQ” pràcticament sempre encerta, permetent rebutjar automàticament mostres inacceptables i estalviar temps en el procés de diagnòstic de malalties del tracte respiratori inferior.

No obstant això, la baixa fidelitat de les màscares i la subrepresentació de la classe BQ limiten la seva aplicació en tasques que requereixin segmentacions molt precises o una cobertura equilibrada de totes les categories. Per això, en aquest cas no es recomana acceptar automàticament les mostres marcades com a “BQ” sense revisió d’un professional a causa de la presència de falsos positius.

Per tal de millorar la robustesa hem considerat recomanable ampliar el conjunt de dades amb més exemples de BQ, incorporar augmentacions específiques per diversificar l’aparença de les classes minoritàries i ajustar els paràmetres de supressió no màxima i de llindars de confiança per reduir falsos positius en BQ. Amb aquestes millores, el model podria assolir un balanç òptim entre exactitud, qualitat de segmentació i velocitat de processament.

Finalment, cal destacar que el model combina velocitat, fiabilitat i adaptabilitat, reforçada per una arquitectura unificada *anchor-free* basada en CSPDDarknet com a *backbone*, un *neck* PAN-FPN i caps desacoblats de classificació i segmentació, juntament amb l’ús de segmentacions avançades (mosaics, CutMix i variacions d’escala i brillantor), posicionant el sistema com una eina potencialment útil per al cribratge ràpid de mostres acceptables i no acceptables en entorns clínics o de laboratori.

DISCUSSION

La valoració de la qualitat de les mostres d'esput mitjançant la tècnica de tinció de Gram és un pas crític i necessari en el diagnòstic d'infeccions respiratòries, ja que determina si una mostra és representativa del tracte inferior i apta per al posterior cultiu microbiològic. La presència abundant de leucòcits polimorfonuclears indica la presència d'una resposta inflamatòria activa i suggereix que hi ha una infecció. Per l'altre costat, un excés de cèl·lules epitelials escamoses indica contaminació provinent de la zona orofaríngea, el que pot comprometre la representació dels resultats. Tot i que es tracta d'un procés clínicament rellevant, la seva interpretació continua sent manual, subjectiva i tediosa. És per això que l'automatització d'aquest procediment mitjançant tècniques d'intel·ligència artificial, tal com la proposta en aquest projecte, ofereix una oportunitat d'augmentar la fiabilitat diagnòstica i d'optimitzar l'ús dels recursos microbiològics.

Aquest estudi tenia com a objectiu explorar la viabilitat tècnica d'un sistema basat en YOLOv8-seg capaç de detectar leucòcits i cèl·lules epitelials per classificar mostres d'esput segons els criteris de Murray i Washington. El model, que s'ha anomenat MurrAI Quality, es va entrenar durant 50 èpoques utilitzant imatges microscòpiques de les mostres d'esput anotades i es va avaluar en dos escenaris diferents.

Els resultats obtinguts mostren que el model MurrAI Quality presenta un rendiment heterogeni en funció de la tasca analitzada. La sensibilitat elevada en la detecció de regions de mala qualitat (MQ), amb un *recall* proper a 0,83, evidencia que l'arquitectura YOLOv8-seg ha après de manera eficaç les textures i patrons presents a aquestes microfotografies. A nivell clínic, aquesta capacitat resulta especialment rellevant en entorns on l'objectiu principal sigui filtrar automàticament preparacions de qualitat deficient, ja que minimitza la taxa de falsos negatius en la detecció de mostres potencialment problemàtiques.

Tanmateix, la precisió en la segmentació fina presenta limitacions importants. El valor de mAP₅₀ no supera 0,23 i es redueix per sota de 0,11 amb un llindar d'IOU de 0,95, indicant una superposició només parcial entre les màscares predites i les

anotacions manuals. Aquest comportament pot atribuir-se, d'una banda, a les diferències de contrast i forma que presenten les cèl·lules en les microfotografies i, de l'altra, a la manca de diversitat d'exemples de bona qualitat (BQ) en el conjunt d'entrenament.

Quan es reagrupen les categories MQ i PQ per plantejar una classificació binària contra BQ, el model assoleix una exactitud del 92% i un *F1-score* ponderat de 0,93. Aquesta millora es deu al fet que la simplificació de la tasca redueix l'impacte de les variacions extremes i incrementa la robustesa del sistema. No obstant això, el suport limitat per a BQ en el conjunt de prova obliga a interpretar aquestes mètriques amb prudència i reforça la necessitat d'ampliar el conjunt de validació per confirmar-ne la generalització.

Des d'un punt de vista operatiu, la latència d'inferència de prop de 3 ms per imatge en GPU posiciona MurrAI Quality com un candidat viable per a la seva integració en entorns clínics o laboratoris d'alt rendiment, on el processament en temps real és un requisit crític. Així mateix, l'ús d'estratègies d'augmentació avançades (mosaic, CutMix i variacions d'escala i brillantor) ha contribuït a millorar la capacitat de generalització davant canvis en la il·luminació i el contrast.

La necessitat d'automatitzar i estandarditzar processos analítics en microbiologia clínica i en tot tipus d'aplicacions biomèdiques és creixent, ja que es busca minimitzar la variabilitat dels resultats i millorar els fluxos de treball. En aquest sentit, diversos estudis han aplicat amb èxit algorismes de detecció i classificació d'imatges biomèdiques, el que evidencia la seva viabilitat i impacte clínic.

Com a exemple, Sun et al. (2023) van desenvolupar un sistema basat en un algorisme YOLOv8 per al cribratge de *Nocardia* en mostres d'esput tenyides per Gram. *Nocardia* és un bacteri oportunista d'importància clínica, especialment en pacients immunodeprimits, i el seu diagnòstic precoç és crític per evitar complicacions greus. No obstant, la seva detecció és complexa, ja que presenta bacils filamentosos i ramificats i, addicionalment, el personal de laboratori no sol estar familiaritzat amb la seva morfologia per la seva raresa. El model desenvolupat va ser avaluat en dos conjunts de dades, donant com a resultat precisions del 97,3%

i 98,3% amb mAP50 de 0,780 i 0,841 i un temps d'inferència de 24 ms. Aquest estudi evidencia que alguns algorismes de detecció poden superar les limitacions humanes en la identificació de morfologies atípiques en entorns clínics.

En una altra aplicació biomèdica, el sistema iMAGING desenvolupat per Maturana et al. (2023) proposa un sistema automatitzat per al diagnòstic de la malària mitjançant un microscopi robotitzat i IA per detectar *Plasmodium* en frotis de sang. La detecció d'aquest paràsit dins dels eritròcits és un procediment essencial per confirmar un diagnòstic, però requereix un examen precís de centenars de camps de la làmina, el que és una tasca lenta i altament subjectiva. Utilitzant un algorisme YOLOv5x, iMAGING va aconseguir una precisió del 92,1%, un *recall* del 93,5%, un *F-score* del 92,79% i mAP50 del 94,4%.

Els exemples amb *Nocardia* i *Plasmodium* mostren la capacitat dels models automatitzats d'identificar estructures morfològicament complexes en entorns heterogenis. Aquest enfocament es pot estendre a àrees com l'oncologia, on és necessari diferenciar lesions tumorals per patrons histològics subtils en teixits molt variables. Rong et al. (2023) va desenvolupar el model HD-YOLO (*Histology-based Detection using YOLO*), capaç de detectar i segmentar nuclis cel·lulars en mostres de pulmó, fetge i mama. Aquest procés no només implica localitzar els nuclis, sinó també diferenciar-los d'altres estructures presents com estroma i vasos sanguinis. La precisió i rapidesa superiors del model, comparat amb els mètodes clàssics, demostren que aquests algorismes poden capturar trets morfològics clau per avaluar processos neoplàsics, on el diagnòstic sovint depèn de variacions molt subtils en mida, forma i distribució nuclear.

En conjunt, aquests avenços evidencien que la integració d'algorismes de detecció en tasques biomèdiques aporta rapidesa, objectivitat i escalabilitat. MurrAI Quality aporta aquesta capacitat a una de les primeres etapes del procés diagnòstic, assegurant que només les mostres realment representatives siguin analitzades. El nostre algorisme constitueix un punt de partida sòlid per a la classificació automàtica de la qualitat de mostres, però cal continuar treballant per millorar la segmentació fina i equilibrar la representació de les classes.

CONCLUSIONS

MurrAI Quality ha demostrat un potencial destacable per a la classificació ràpida i automatitzada de la qualitat de microfotografies d'esput, aconseguint una exactitud superior al 90% en l'escenari binari que distingeix entre mostres acceptables (MQ+PQ) i preparacions de bona qualitat (BQ). La seva baixa latència d'inferència permet plantejar la seva integració en entorns clínics d'alt rendiment, mentre que el *recall* elevat en la detecció de mala qualitat garanteix una mínima omisió de mostres crítiques. Això demostra que, malgrat les limitacions en la segmentació fina, el sistema és efectiu quan l'objectiu sigui identificar ràpidament les mostres de mala qualitat, evitant així perdre temps i recursos en el seu processament.

Tot i això, la precisió de les màscares segmentades continua sent limitada, amb un mAP50 inferior al 25% i una disminució pronunciada en augmentar el llindar IoU. Aquest resultat subratlla la necessitat d'enriquir el conjunt de dades amb més exemples de preparacions de bona qualitat, així com d'explorar funcions de pèrdua de segmentació més especialitzades.

La subrepresentació de la classe BQ en el conjunt de validació introdueix una certa inestabilitat en les mètriques obtingudes. Per tant, les línies de treball futures haurien d'orientar-se a equilibrar la distribució de classes mitjançant noves anotacions i tècniques d'augmentació dirigides, a més d'optimitzar els llindars i el postprocessat per reduir falsos positius.

Amb aquestes millores, l'arquitectura YOLOv8-seg emprada per MurrAI Quality podria evolucionar des d'un prototip prometedora a una eina robusta i fiable per al control automàtic de qualitat en imatge mèdica, combinant la velocitat del processament en temps real amb una segmentació de major precisió.

AUTOAVALUACIÓ

Durant el desenvolupament d'aquest TFG he tingut l'oportunitat d'aprendre tant de l'àmbit de la microbiologia com de la intel·ligència artificial. Per una banda, la part del treball al laboratori amb els professionals del Servei de Microbiologia del Centre de Diagnòstic Biomèdic de l'Hospital Clínic m'ha permès aprendre la importància de l'avaluació acurada de les mostres i de la necessitat d'estandarditzar aquest procediment per un millor diagnòstic clínic. D'altra banda, ha suposat un repte constant, ja que al principi desconeixia les eines i conceptes tècnics relacionats amb les xarxes neuronals convolucionals, el processament d'imatges i l'entrenament amb models de *deep learning*.

Pel que fa als resultats, estic satisfeta amb el que hem aconseguit. Tot i que hi ha encara molt marge de millora considero que les limitacions són coherents amb el temps i els recursos disponibles i que aquest estudi és un bon punt de partida per a futurs desenvolupaments.

En conjunt, considero que l'experiència ha estat molt enriquidora tant en l'àmbit acadèmic com el personal. M'ha permès adquirir competències clau per a la meva futura vida laboral en la biotecnologia que no havia tingut l'oportunitat de desenvolupar durant la carrera. Alhora, m'ha despertat un gran interès en continuar formant-me en aquest camp donada la seva projecció i les nombroses aplicacions que pot tenir en el futur.

BIBLIOGRAFIA

- Aung, Y. Y. M., Wong, D. C. S., & Ting, D. S. W. (2021). The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare. In *British Medical Bulletin* (Vol. 139, Issue 1, pp. 4–15). Oxford University Press. <https://doi.org/10.1093/bmb/ldab016>
- Bender, R. G., Sirota, S. B., Swetschinski, L. R., Dominguez, R. M. V., Novotney, A., Wool, E. E., Ikuta, K. S., Vongpradith, A., Rogowski, E. L. B., Doxey, M., Troeger, C. E., Albertson, S. B., Ma, J., He, J., Maass, K. L., Simões, E. A. F., Abdoun, M., Aziz, J. M. A., Abdulah, D. M., ... Kyu, H. H. (2024). Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990–2021: a systematic analysis from the Global Burden of Disease Study 2021. *The Lancet Infectious Diseases*, 24(9), 974–1002. [https://doi.org/10.1016/S1473-3099\(24\)00176-2](https://doi.org/10.1016/S1473-3099(24)00176-2)
- Bouzada, F. M., Mestre, B., Vaquer, A., Tejada, S., & de la Rica, R. (2025). Detecting Respiratory Pathogens for Diagnosing Lower Respiratory Tract Infections at the Point of Care: Challenges and Opportunities. In *Biosensors* (Vol. 15, Issue 3). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/bios15030129>
- Budayanti, N. S., Suryawan, K., Iswari, I. S., & Sukrama, D. M. (2019). The quality of sputum specimens as a predictor of isolated bacteria from patients with lower respiratory tract infections at a Tertiary Referral Hospital, Denpasar, Bali-Indonesia. *Frontiers in Medicine*, 6(APR). <https://doi.org/10.3389/fmed.2019.00064>
- Cacho, J. B., María, C., Meseguer, A., Antonio, P., Palomo, O., & Puig De La Bellacasa, J. (s.d.). Editors: Emilia Cercenado y Rafael Cantón Coordinadora: María Antonia Meseguer Peinado Autores.

Cartuliales, M. B., Skjøt-Arkil, H., Mogensen, C. B., Skovsted, T. A., Andersen, S. L., Pedersen, A. K., & Rosenvinge, F. S. (2023). Gram Stain and Culture of Sputum Samples Detect Only Few Pathogens in Community-Acquired Lower Respiratory Tract Infections: Secondary Analysis of a Randomized Controlled Trial. *Diagnostics*, 13(4). <https://doi.org/10.3390/diagnostics13040628>

Centre de Diagnòstic Biomèdic | Hospital Clínic Barcelona. (2025, 1 agost). Clínic Barcelona.

<https://www.clinicbarcelona.org/ca/institut/centre-de-diagnostic-biomedic>

Contributing Writer. (May 22, 2025). Data Annotation for High-Performing Computer Vision Models. Roboflow Blog: <https://blog.roboflow.com/data-annotation/>

Gary W. Procop, D. L. C. G. S. H. W. M. J. (2006). *Koneman's Color Atlas And Textbook Of Diagnostic Microbiology* (6th ed.). Lippincott Williams & Wilkins. <https://archive.org/details/konemanscoloratl0000unse>

Haifa Alanazi. (2025). Optimizing Medical Image Analysis: A Performance Evaluation of YOLO-Based Segmentation Models. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 16(4), 1167–1174.

Jeczminek, E., & Kowalski, P. A. (2021). Flattening Layer Pruning in Convolutional Neural Networks. *Symmetry*, 13(7), 1147. <https://doi.org/10.3390/sym13071147>

Johny, F., & Thomas S, K. (2024). Sputum Quality Assessment as a Useful Tool for Distinguishing True Pathogens from Colonizing Flora. *RGUHS Journal of Allied Health Sciences*, 4(3). https://doi.org/10.26463/rjahs.4_3_5

Kenneth P. Smith, H. W. T. J. S. D. B. A. M. S. E. S. J. E. K. S. W. L. D. D. R. (1995). Clinical microbiology newsletter. *Clinical Microbiology Newsletter*, 42(8), 61–70.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. In *Nature* (Vol. 521, Issue 7553, pp. 436–444). Nature Publishing Group. <https://doi.org/10.1038/nature14539>

Maturana, C. R., de Oliveira, A. D., Nadal, S., Serrat, F. Z., Sulleiro, E., Ruiz, E., Bilalli, B., Veiga, A., Espasa, M., Abelló, A., Suñé, T. P., Segú, M., López-Codina, D., Clols, E. S., & Joseph-Munné, J. (2023). iMAGING: a novel automated system for malaria diagnosis by using artificial intelligence tools and a universal low-cost robotized microscope. *Frontiers in Microbiology*, 14. <https://doi.org/10.3389/fmicb.2023.1240936>

Murel, J., PhD, & Kavlakoglu, E. (2024, 3 gener). What is object detection? *IBM*. <https://www.ibm.com/think/topics/object-detection>

Ogawa, H., Kitsios, G. D., Iwata, M., & Terasawa, T. (2020). Sputum Gram Stain for Bacterial Pathogen Diagnosis in Community-acquired Pneumonia: A Systematic Review and Bayesian Meta-analysis of Diagnostic Accuracy and Yield. *Clinical Infectious Diseases*, 71(3), 499–513. <https://doi.org/10.1093/cid/ciz876>

Pratiher, D. (2024, 12 maig). Understanding loss functions for deep learning segmentation models. Medium. <https://medium.com/@devanshipratiher/understanding-loss-functions-for-deep-learning-segmentation-models-30187836b30a>

Ragab, M. G., Abdulkadir, S. J., Muneer, A., Alqushaibi, A., Sumiea, E. H., Qureshi, R., Al-Selwi, S. M., & Alhussian, H. (2024). A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). *IEEE Access*, 12, 57815–57836. <https://doi.org/10.1109/ACCESS.2024.3386826>

Rong, R., Sheng, H., Jin, K. W., Wu, F., Luo, D., Wen, Z., Tang, C., Yang, D. M., Jia, L., Amgad, M., Cooper, L. A. D., Xie, Y., Zhan, X., Wang, S., & Xiao, G.

- (2023). A Deep Learning Approach for Histology-Based Nucleus Segmentation and Tumor Microenvironment Characterization. *Modern Pathology*, 36(8), 100196. <https://doi.org/10.1016/j.modpat.2023.100196>
- Shah, D. (s. d.). Intersection over Union (IoU): Definition, Calculation, Code. V7. <https://www.v7labs.com/blog/intersection-over-union-guide>
- Shah, D. (s. d.). Mean Average Precision (mAP) Explained: Everything You Need to Know. V7. <https://www.v7labs.com/blog/mean-average-precision>
- Shen F, Zubair M, Sergi C. Sputum Analysis. [Updated 2025 May 4]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK563195/>
- Sirohi, M., Lall, M., Yenishetti, S., Panat, L., & Kumar, A. (2022). Development of a Machine learning image segmentation-based algorithm for the determination of the adequacy of Gram-stained sputum smear images. *Medical Journal Armed Forces India*, 78(3), 339–344. <https://doi.org/10.1016/j.mjafi.2021.09.012>
- Smith, K. P., Wang, H., Durant, T. J. S., Mathison, B. A., Sharp, S. E., Kirby, J. E., Long, S. W., & Rhoads, D. D. (2020). Applications of Artificial Intelligence in Clinical Microbiology Diagnostic Testing. *Clinical Microbiology Newsletter*, 42(8), 61–70. <https://doi.org/10.1016/j.clinmicnews.2020.03.006>
- Sun, H., Xie, X., Wang, Y., Wang, J., & Deng, T. (2023). Clinical screening of *Nocardia* in sputum smears based on neural networks. *Frontiers in Cellular and Infection Microbiology*, 13. <https://doi.org/10.3389/fcimb.2023.1270289>
- Torres, J. (2025, agosto 1). YOLOv8 Architecture; Deep Dive into its Architecture Yolov8. YOLOv8. <https://yolov8.org/yolov8-architecture/#Introduction>

Torres, J. (2025, gener 3). YOLOV8 Segmentation; Step-by-step explained - Yolov8. YOLOv8.

https://yolov8.org/yolov8-segmentation/#What_is_YOLOv8_Segmentation

Torres, J. (2025, gener 3). YOLOv8 Data Augmentation: Secret to Success | YOLOv8. YOLOv8.

https://yolov8.org/yolov8-data-augmentation/#What_is_YOLOv8_Data_Augmentation

Ultralytics. (2025, 26 juny). YOLO Performance Metrics.

<https://docs.ultralytics.com/es/guides/yolo-performance-metrics/>