

Degree in Data Science and Engineering

Title: Breast tumor detection in Hematoxylin & Eosin stained biopsy images

Author: Sonia Rabanaque Rodríguez

Advisor: Josep Ramon Casas & Verónica Vilaplana

Department: Grup de Processament d'Imatge i Vídeo

Month and year: June, 2021



**UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH**

**Facultat d'Informàtica de Barcelona
Facultat de Matemàtiques i Estadística
Escola Tècnica Superior d'Enginyeria de Telecomunicació de
Barcelona**

CATALÀ

El càncer fa referència a un conjunt de malalties molt presents en la nostra societat. Gràcies als nous avenços en el camp de la medicina, la taxa de supervivència en els diferents tipus de càncer augmenta contínuament. Un dels aspectes fonamentals és fer una detecció ràpida de la malaltia mitjançant diverses proves diagnòstiques. En el cas del càncer de mama, s'extreu una mostra de teixit a partir d'una biòpsia, es tenyeix amb una tinció anomenada Hematoxilina-Eosina i s'analitza per tal de detectar regions amb cèl·lules tumorals. Aquesta tasca és molt exigent degut al gran volum de feina que comporta, ja que els patòlegs han d'analitzar tota la mostra de teixit escanejada en imatges histològiques digitals. Són imatges d'una mida molt gran degut al gran nivell de detall amb el que s'ha capturat la mostra.

Aquest treball es desenvolupa dins del projecte DigiPatICS, que té com a objectiu proporcionar algorismes d'anàlisi per l'Institut Català de la Salut (ICS). Així doncs, per ajudar als patòlegs en la detecció del càncer de mama, l'objectiu d'aquest treball és entrenar una xarxa neuronal per tal d'identificar les regions tumorals presents en el teixit. Concretament, s'entrena una arquitectura U-Net per tal de fer la segmentació de les imatges en 5 classes diferents, entre elles, tumor. Degut a la impossibilitat d'obtenir imatges anotades pròpies per aquest projecte, s'han utilitzat les del *Breast Cancer Semantic Segmentation (BCSS) challenge*. Per tant, en aquest projecte s'entrena la xarxa neuronal amb aquestes dades i, un cop obtingut el model òptim, es fa inferència utilitzant les dades de l'Institut Català de la Salut.

El model s'ha entrenat a partir de les dades de BCSS amb les quals s'han obtingut uns resultats molt bons, sobretot en la classe referent al tumor. Pel que fa la inferència, les prediccions obtingudes també són força bones, tot i que no ha sigut possible mesurar els resultats quantitativament per la manca d'anotacions o bé obtenir impressions per part dels patòlegs dels resultats obtinguts. Així doncs, s'espera en un futur poder validar i/o corregir aquestes prediccions de les imatges de l'ICS per tal de poder re-entrenar la xarxa i millorar el seu funcionament.

CASTELLÀ

El càncer hace referencia a un conjunto de enfermedades muy presentes en nuestra sociedad. Gracias a los nuevos avances en el campo de la medicina, la tasa de supervivencia en los diferentes tipos de cáncer aumenta continuamente. Uno de los aspectos fundamentales es hacer una detección rápida de la enfermedad mediante diversas pruebas diagnósticas. En el caso del cáncer de mama, se extrae una muestra de tejido a partir de una biopsia, se tiñe con una tinción llamada Hematoxilina-Eosina y se analiza para detectar regiones con células tumorales. Esta tarea es muy exigente debido al gran volumen de trabajo que conlleva, ya que los patólogos tienen que analizar toda la muestra de tejido escaneada en imágenes histológicas digitales. Son imágenes de un tamaño muy grande debido al gran nivel de detalle con el que se ha capturado la muestra.

Este trabajo se desarrolla dentro del proyecto DigiPatICS, cuyo objetivo es proporcionar algoritmos de análisis para el Instituto Catalán de la Salud (ICS). Así que, para ayudar a los patólogos en la detección del cáncer de mama, el objetivo de este trabajo es entrenar una red neuronal para identificar las regiones tumorales presentes en el tejido. Concretamente, se entrena una arquitectura U-Net para hacer la segmentación de las imágenes en 5 clases diferentes, entre ellas, tumor. Debido a la imposibilidad de obtener

imágenes anotadas propias para este proyecto, se han utilizado las del *Breast Cancer Semantic Segmentation (BCSS) challenge*. Por lo tanto, en este proyecto se entrena la red neuronal con estos datos y, una vez obtenido el modelo óptimo, se hace inferencia utilizando los datos del Instituto Catalán de la Salud.

El modelo se ha entrenado a partir de los datos de BCSS con los que se han obtenido unos resultados muy buenos, sobre todo en la clase referente al tumor. En cuanto a la inferencia, las predicciones obtenidas también son bastante buenas, aunque no ha sido posible medir los resultados cuantitativos por la falta de anotaciones, ni la impresión por parte de los patólogos sobre los resultados obtenidos. Así que se espera en un futuro poder validar y/o corregir estas predicciones de las imágenes del ICS para poder reentrenar la red y mejorar su funcionamiento.

ANGLÈS

Cancer refers to a group of diseases that are very present in our society. Thanks to new advances in the field of medicine, the survival rate in different types of cancer is continuously increasing. One of the fundamental aspects is to make a fast detection of the disease through various diagnostic tests. In the case of breast cancer, a tissue sample is extracted from a biopsy, stained with a stain called Hematoxylin-Eosin, and analysed for regions with tumour cells. This task is very tough due to the large volume of work involved, as pathologists have to analyse the entire scanned tissue sample on digital histological images. These images are very large due to the big level of detail with which the sample has been captured.

This work is developed within the DigiPatICS project, which aims at providing algorithms for the Catalan Institute of Health (ICS). So, to help pathologists in the detection of breast cancer, the objective of this work is to train a neural network to identify the tumour regions present in the tissue. Specifically, a U-Net architecture is trained to segment the images into 5 different classes, including tumour. Due to the impossibility of obtaining our own annotated images for this project, those of the Breast Cancer Semantic Segmentation (BCSS) challenge have been used. Therefore, in this project the neural network is trained with these data and, once the optimal model is obtained, the inference is made using data from the Catalan Health Institute.

Therefore, the model has been trained from the BCSS data with which very good results have been obtained, especially in the tumour class. In terms of inference, the predictions obtained are also quite good, although it has not been possible to obtain quantitative results due to the lack of annotations or getting impressions from the pathologists on the results obtained. So, we hope that in the future it can be able to validate and/or correct these predictions of the ICS images in order to retrain the network and improve its performance.

Índex

1.	Introducció i context del problema	2
2.	Definició, objectius i especificacions del problema	3
3.	Treball previ i fonaments teòrics	5
3.1.	Treball previ.....	5
3.2.	Aspectes mèdics.....	7
3.2.1.	Càncer de mama	7
3.2.2.	Proves diagnòstiques.....	9
3.2.3.	Whole Slide Images (WSI)	11
3.3.	Aspectes tecnològics	12
3.3.1.	Arquitectura CNN.....	12
3.3.2.	Breast Cancer Semantic Segmentation (BCSS) dataset	20
3.3.3.	Dades de l'ICS	21
4.	Solució proposada.....	23
4.1.	BCSS dataset.....	24
4.2.	Arquitectura CNN.....	26
4.3.	Preparació dades de l'ICS	27
4.3.1.	Creació dels <i>patches</i>	27
4.3.2.	Eina de visualització.....	29
4.4.	Registre.....	29
5.	Resultats	32
5.1.	Resultats del BCSS dataset.....	32
5.2.	Inferència dades de l'ICS	39
5.3.	Registre.....	41
6.	Conclusions i treball futur	43
7.	Bibliografia	45

1. Introducció i context del problema

El càncer de mama fa referència a un tumor maligne que s'ha desenvolupat a partir de cèl·lules mamàries degut al seu creixement descontrolat. És un dels tumors més coneguts, ja que és la primera causa de morts per càncer en dones. Per exemple, l'any 2019, es van detectar 33.307 nous casos de càncer de mama a Espanya, i aquest número continua creixent lentament. Tot i això, gràcies a les noves teràpies i als nous avenços, tant mèdics com de processament d'imatges, la taxa de supervivència està augmentant i per aquest motiu, per exemple a Espanya, la mortalitat del càncer de mama és de les més baixes.

Per detectar si un pacient té càncer de mama o no, es fa una biòpsia per tal d'extreure una mostra de teixit a analitzar posteriorment. Concretament, s'obtenen diferents talls que s'escanegen per tal d'obtenir unes imatges digitals anomenades *Whole Slide Images*, que són amb les que es treballa en aquest projecte. Els patòlegs utilitzen eines de visualització especialitzades per analitzar el teixit en detall i poder identificar possibles zones tumorals.

Així doncs, l'objectiu d'aquest projecte és implementar una eina basada en xarxes neuronals convolucionals que ajudi als especialistes en la tasca de la detecció de zones tumorals a partir de les imatges del teixit. D'aquesta manera, els especialistes tindran l'ajuda d'una eina que els permeti treballar de manera més ràpida i precisa a l'hora de fer els diagnòstics dels diferents pacients. Això és molt important, ja que un dels aspectes clau és poder fer la detecció el més aviat possible i decidir correctament el tractament del pacient.

Aquest projecte es desenvolupa en el Grup de Processament d'Imatge (GPI) de la Universitat Politècnica de Catalunya (UPC). Concretament, forma part d'un projecte més gran anomenat DigiPatICS que s'encarrega de proporcionar algorismes per diferents tasques a l'Institut Català de la Salut (ICS).

L'objectiu del projecte DigiPatICS és desenvolupar una eina que detecti zones tumorals i calculi mètriques per tal de fer la classificació del grau del tumor i, d'aquesta manera, ajudar als patòlegs a fer diagnòstics més ràpids i precisos. Per tal de fer això, la idea és entrenar diferents xarxes neuronals a partir d'anotacions fetes a nivell de cèl·lula o de *patch*, és a dir, a nivell de píxel o de la imatge sencera representat una regió concreta.

Aquest treball, per tant, s'especialitza en una de les tasques requerides per l'ICS. Concretament, s'ha treballat amb una de les tincions utilitzada pels patòlegs, anomenada Hematoxilina-Eosina (HE), que és la que s'utilitza per indicar la presència de zones tumorals. D'aquesta manera, un cop s'han detectat cèl·lules tumorals, es requereixen altres tincions per fer diferents càlculs i obtenir informació útil per fer un diagnòstic més precís, per exemple, sobre el grau del tumor.

2. Definició, objectius i especificacions del problema

Aquest projecte es desenvolupa per tal de detectar tumors de càncer de mama en imatges histològiques utilitzant una xarxa neuronal. Específicament, aquesta detecció es duu a terme en imatges de biòpsies tenyides amb Hematoxilina i Eosina (HE). D'aquesta manera, els patòlegs poden saber quines àrees tenen més probabilitats de ser tumorals i, després, decidir aplicar altres tincions per assegurar-se i conèixer el grau del tumor.

Com s'ha explicat anteriorment, aquest projecte es duu a terme conjuntament amb l'Institut Català de la Salut (ICS), que és el que proporciona tant les imatges a utilitzar com la informació mèdica rellevant per tal de concebre els algorismes.

Els objectius principals del projecte són:

1. Desenvolupar eines d'anàlisi d'imatges i aprenentatge automàtic per ajudar als patòlegs en la detecció de zona tumoral a partir de les *Whole Slide Images*.
2. Desenvolupar eines que funcionin en un entorn real, és a dir, que puguin integrar-se al programari existent.
3. Desenvolupar algorismes que proporcionin retroalimentació, en altres paraules, que puguin aprendre informació a partir de noves dades.
4. Crear eines eficients, que requereixin temps d'execució i recursos computacionals raonables.

Respecte el primer punt, com ja s'ha esmentat anteriorment, la idea és entrenar una xarxa neuronal que aprengui a identificar les zones on és més probable que hi hagi cèl·lules tumorals. Aquesta identificació ha de fer-se en tota la imatge sencera, és a dir, en totes les zones de la imatge que contenen teixit. Per tant, els resultats obtinguts per cadascun dels *patches* s'hauran de "fusionar" per tal d'obtenir el resultat final de manera que sigui útil pels patòlegs. També s'ha de tenir present que el diagnòstic no es fa només a partir d'aquesta tinció, sinó que s'utilitza també la informació present en altres. Per aquest motiu, també és interessant poder combinar la informació complementària de les diferents tincions per obtenir un anàlisi més precís i complet.

A més, és important que aquestes eines puguin funcionar en un entorn de treball real, tal i com s'indica al següent punt. Per tant, s'ha de tenir en compte el programari disponible pels patòlegs i els seus equips a l'hora de programar els algorismes.

Un altre aspecte fonamental és que la xarxa pugui aprendre de noves anotacions fetes pels patòlegs. D'aquesta manera, serà capaç de ser més acurada a l'hora de fer les prediccions i, per tant, de detectar els diferents tipus de tumor. A més, també és important poder afegir o treure classes a la segmentació segons el que els patòlegs creguin oportú. Aquest punt també implica que la xarxa pugui aprendre a predir millor tenint en compte les característiques de les imatges histològiques de l'ICS.

Per últim, la eficiència és una aspecte fonamental en l'eina a desenvolupar. Això implica utilitzar una quantitat de recursos computacionals raonable i que el temps d'execució no sigui massa gran. El motiu és que l'algorisme es pugui executar amb les imatges noves per tal de tenir totes les prediccions el més aviat possible i, d'aquesta manera, s'espera poder ajudar als patòlegs a fer els diagnòstics de manera més eficient.

Aquest projecte també té una sèrie de problemes que s'han de tenir en compte a l'hora d'assolir els objectius especificats anteriorment. Els principals són dos: la dificultat d'aconseguir inicialment dades anotades i la intervariabilitat de les *Whole Slide Images* (WSI) entre els diferents centres de l'ICS.

Respecte el primer punt, la manca d'imatges anotades, s'ha de tenir en compte que crear el *ground truth* suposa una feina molt gran pels patòlegs, tant per l'exigència com pel temps dedicat, ja que aquestes imatges tenen una gran resolució i, per tant, molt detall. Així doncs, és molt costós crear una base de dades anotades per les diferents tasques del projecte DigiPatICS. El segon aspecte, en canvi, està més relacionat amb la preparació de les imatges histològiques, ja que diferents procediments, productes químics i eines poden produir diferents tipus d'imatges. Per tant, l'aparença del teixit pot ser molt diferent i això pot tenir efectes en la predicció per part del model entrenat.

En el cas d'aquest treball en concret, no s'ha pogut disposar d'imatges anotades pels patòlegs de l'ICS. Per tant, s'han cercat diferents bases de dades i s'ha treballat amb la que s'ha cregut més convenient, tant pels membres del grup de DigiPatICS com per alguns metges de l'ICS als quals se'ls hi ha consultat. També s'ha de destacar que, en el moment de la realització d'aquest treball, només s'ha disposat d'imatges de l'Hospital Vall d'Hebron, així que no s'ha pogut observar l'efecte de la intervariabilitat en les diferents imatges dels centres pertanyents a l'ICS.

3. Treball previ i fonaments teòrics

En aquesta secció s'expliquen diferents aspectes que s'han de conèixer per entendre els següents apartats. Primer, es fa una explicació del treball previ sobre aquesta tasca, la detecció de zones tumorals en imatges histològiques. A continuació, s'expliquen conceptes tant mèdics com tecnològics que són d'utilitat per comprendre les decisions preses en aquest treball. Per tal d'entendre la malaltia, es fa una explicació de què és el càncer de mama i el procés complet de detecció. A més, s'explica una mica com és l'estructura de les Whole Slide Images (WSI), el tipus de dades amb les que s'ha treballat. Respecte la part de la implementació, primer es mostra l'arquitectura utilitzada (U-Net) i les diferents mètriques per tal d'avaluar el model i els resultats obtinguts. Finalment, es fa una explicació dels dos conjunts d'imatges que s'han fet servir: el BCSS dataset per entrenar el model, i les imatges de l'ICS que són en les que es vol utilitzar l'algorisme.

3.1. Treball previ

En els últims anys s'han aconseguit grans avenços amb la millora en els àmbits de la medicina i la tecnologia per tal de millorar la qualitat de vida de les persones i reduir les morts produïdes per certes malalties com, per exemple, el càncer. Com ja s'ha esmentat anteriorment, és molt important fer una detecció el més aviat possible i acurada per tal de millorar el diagnòstic i les taxes de supervivència dels pacients.

La detecció de càncer de mama és una tasca que requereix molt de temps i que és exigent, ja que implica que els patòlegs han d'examinar grans trossos de regió benigna fins identificar les àrees de malignitat. En aquest sentit, els sistemes de diagnòstic assistits per ordinadors contribueixen a reduir el cost i a augmentar l'eficiència del diagnòstic.

Les primeres tècniques que s'utilitzaven per a la detecció de cèl·lules cancerígenes es basaven en l'extracció de característiques per tal d'utilitzar un classificador (*Support Vector Machines*, *random forests*, etc.) per fer la distinció entre zones benignes i malignes. Un dels problemes d'aquests mètodes és que estan dissenyats per un problema específic i, per tant, no generalitzen bé per altres tasques semblants.

En canvi, les arquitectures de *deep learning* han mostrat tenir molt èxit en la segmentació automàtica i en la classificació en imatges histològiques digitalitzades. Amb les xarxes neuronals es poden obtenir representacions més abstractes i útils sobre les dades. D'aquesta manera, es poden aprendre patrons amagats que no poden ser perceptibles pels patòlegs ni per les tècniques basades en l'extracció de característiques.

Per tal de detectar tumors de càncer de mama en imatges histològiques utilitzant una xarxa neuronal s'han considerat dos enfocaments diferents: classificació de *patches* i segmentació semàntica. El primer cas consisteix en decidir la classe de cadascun dels *patches* en els que es divideix la imatge, mentre que, en la segmentació, per cadascuna d'aquestes regions, el resultat és una classificació a nivell de píxel per tal d'obtenir una detecció precisa de la zona tumoral.

Actualment, la majoria de projectes realitzats amb *Whole Slide Images* (WSI) de tincions d'Hematoxilina i Eosina (HE) es centren en fer classificació de *patches*. Al 2014, es va publicar *Automatic detection of invasive ductal carcinoma in whole slide images with Convolutional Neural Networks*. En aquest, Cruz-Roa *et al.* [1] entrenen una xarxa convolucional (CNN) amb l'objectiu de detectar regions amb carcinoma ductal invasiu (IDC, sigles en anglès). Concretament, es classifiquen *patches* segons si contenen aquest tipus de tumor o no.

Més tard, l'any 2017, el treball d'Araújo *et al.* [2] només utilitza imatges de càncer de mama. La idea és fer una classificació de la imatge sencera, fent primer una classificació a nivell de *patch* i després combinant els resultats per obtenir la classe a nivell d'imatge. En aquest treball s'utilitza una CNN tant per extreure les *features* com per fer la classificació en quatre classes: teixit normal, lesió benigna, carcinoma *in situ* i carcinoma invasiu.

A continuació, Gecer *et al.* [3], mitjançant una localització multi-escala de les regions d'interès, fa una classificació en cinc classes: canvis proliferatius, canvis no proliferatius, hiperplàsia ductal atípica, carcinoma ductal *in situ* i carcinoma invasiu. Primer, fan servir una *Fully Convolutional Network* (FCN) per localitzar regions d'interès i, després, utilitzen una CNN per identificar les cinc categories mencionades anteriorment.

El mateix any es va publicar un treball molt complet [4] en el qual s'estudien diferents problemes relacionats amb l'anàlisi d'imatges histològiques com classificació de limfomes, detecció de Carcinoma Ductal Invasiu (IDC), segmentació de nuclis, etc. En el cas de la detecció d'IDC, utilitzen una *Inception Recurrent Residual Convolutional Neural Network* (IRRCNN) per classificar els *patches* segons si contenen un carcinoma d'aquest tipus o no.

Per últim, el més recent relacionat amb la classificació és el treball d'Iizuka *et al.* [5], publicat al 2020 i que treballa amb imatges d'altres tipus de tumors. Els models amb els que treballen són CNNs i *Recurrent Neural Networks* (RNNs) per tal de fer una classificació de *patches* i de la WSI sencera, respectivament.

En canvi, hi ha menys treballs publicats relacionats amb la segmentació d'imatges histològiques d'aquest tipus entrenant xarxes neuronals. Hi ha alguns en els quals s'utilitzen altres mètodes com en el cas de Qu *et al.* [6], on primer s'extreuen *features* (a partir del color i la textura) i es construeixen classificadors a nivell de píxel utilitzant *Support Vector Machines* per tal de fer la segmentació. El resultat final s'obté després de fer un post-processament per evitar problemes de soroll.

Un dels treballs que sí que entrena una xarxa neuronal és el publicat per Guo *et al.* [7], de l'any 2019. El propòsit d'aquest treball és diferent als explicats anteriorment, ja que l'objectiu és fer una segmentació. Primer, s'utilitza un model *Inception v3* per tal de fer una pre-selecció de les regions amb tumor i, a continuació, es fa servir un model DCNN (*Deep Convolutional Neural Network*) per fer la segmentació semàntica final.

Aresta *et al.* [8] presenten algunes de les arquitectures utilitzades en el BACH challenge. Aquestes són Convolutional Neural Networks (CNN) en la majoria dels casos utilitzant Inception v3, ResNext i DenseNet.

A partir d'aquests treballs previs, es pot concloure que les CNNs obtenen bons resultats en aplicacions patològiques com la classificació i segmentació de tumor, que és l'objectiu d'aquest treball. No només en aquest tipus de tasques, sinó que també han

demostrat bons resultats i beneficis en una gran quantitat d'aplicacions relacionades amb *computer vision* i anàlisi d'imatges mèdiques. Per tant, semblen bons candidats per desenvolupar eines per ajudar als patòlegs en la millora del diagnòstic, tant en la precisió com en l'eficiència.

3.2. Aspectes mèdics

En aquesta subsecció i les següents introduïm diversos aspectes mèdics relacionats amb el càncer de mama, el seu diagnòstic i les imatges obtingudes a partir de la mostra de teixit.

3.2.1. Càncer de mama

El càncer [9], [10] és una malaltia que es caracteritza pel desenvolupament de cèl·lules anormals i que es produeix degut a les mutacions o canvis anòmals en els gens que regulen el creixement de les cèl·lules i que les mantenen sanes. Aquestes mutacions poden "activar" alguns gens i "desactivar"-ne d'altres. Mentre que les cèl·lules normals es divideixen i moren durant un temps programat, les cèl·lules canceroses "perden" la capacitat de morir i es divideixen quasi sense límits, de manera que produeix més cèl·lules iguals i generen tumors, que en la seva expansió poden destruir i substituir als teixits normals.

Els tumors poden ser malignes o benignes, segons si són perillosos o no per a la salut, respectivament. Aquests últims no es consideren cancerosos, ja que les seves cèl·lules tenen una aparença relativament normal, creixen lentament i no envaeixen teixits pròxims ni es propaguen cap a altres parts del cos. Per altra banda, els tumors malignes sí són cancerosos i, de no ser controlats, les seves cèl·lules poden propagar-se cap a altres parts del cos. Si es dóna aquest cas, es parla de metàstasi.

A partir de l'anterior, es pot dir que el càncer de mama [10], [11] fa referència a un tumor maligne que s'ha desenvolupat a partir de cèl·lules mamàries degut al seu creixement descontrolat. Aquest tipus de càncer es pot produir tant en homes com en dones, però és molt més comú en aquestes últimes.

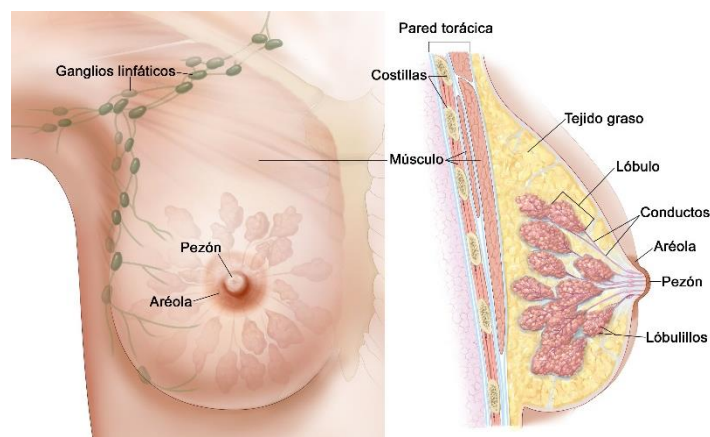


Figura 1. Anatomia de la mama.

Generalment, aquest tipus de càncer s'origina als lòbuls (glàndules que produeixen la llet) o als conductes (vies que transporten la llet fins als mugrons), representats a la

Figura 1. En aquests casos, es parla de carcinoma lobular i carcinoma ductal, respectivament. Tot i així, també es pot produir a altres tipus de cèl·lules o teixit dins la mama. Aquestes cèl·lules poden arribar al teixit mamari sa del voltant i arribar als ganglis limfàtics (petits òrgans que eliminen substàncies estranyes de l'organisme) que es troben a les aixelles. A partir d'aquests, poden propagar-se per la resta del cos (metàstasi).

El càncer de mama s'origina sempre per una anomalia genètica, és a dir, per un "error" en el material genètic. Tot i això, només un 5 – 10% dels casos són deguts a mutacions genètiques heretades. En canvi, el 85 – 90% dels casos tenen el seu origen en anomalies vinculades a l'envelliment.

Degut a l'augment de la conscienciació i a la investigació, s'han aconseguit crear nous avenços tant en el diagnòstic com en el tractament d'aquesta malaltia. Degut a això, la taxa de supervivència ha augmentat i el número de morts associades a aquest tipus de càncer està disminuint constantment, ja sigui per la detecció precoç, el nou enfoc de tractaments personalitzats i a que es té més coneixement de la malaltia.

Clàssicament, l'estudi del càncer estava centrat en la cèl·lula tumoral, però en l'actualitat existeixen proves sòlides que el desenvolupament del càncer i la seva progressió depenen d'una interacció complexa del tumor amb altres elements. Per aquest motiu, en els últims anys, l'estudi del càncer s'amplia al microambient tumoral [12], és a dir, al conjunt de cèl·lules i mecanismes tissulars que poden intervenir en el desenvolupament i en la formació del tumor. Per tant, un tumor no és simplement un grup de cèl·lules canceroses, sinó més aviat una entitat complexa i en contínua evolució. Els diferents elements que formen o que poden estar presents en aquest microambient tumoral són el tumor, l'estroma, la infiltració inflamatòria i la necrosi, entre d'altres.

TUMOR

Un tumor és un conjunt de cèl·lules de qualsevol teixit que s'han multiplicat descontroladament, formant una massa anormal. Existeixen dos tipus de tumors: benignes i malignes.

Les mames contenen una xarxa de vasos sanguinis (sistema vascular) i de canals limfàtics (sistema limfàtic) que transporten sang i líquids d'anada i de tornada des del teixit mamari fins a la resta del cos. Són una espècie d'"autopistes" que aporten aliment i eliminen la sang usada i les deixalles de la vida cel·lular. Hi ha diferents tipus de tumors en el cas del càncer de mama, però els dos principals són els següents:

- Angioinvasió [13]: És la disseminació del tumor cap a un vas sanguini. Per tant, la invasió de sistema vascular es produeix quan les cèl·lules canceroses irrompen en els vasos sanguinis. Això augmenta el risc que el càncer s'estengui fora de la mama o torni en el futur.
- DCIS [14]: El carcinoma ductal in situ (DCIS) és la presència de cèl·lules anormals en el conducte mamari. Això significa que les cèl·lules que revesteixen els conductes per on circula la llet s'han convertit en canceroses, però no s'han propagat al teixit mamari circumdant. Es considera un càncer no invasiu o preinvasiu, però s'ha de tractar perquè pot convertir-se en invasiu.

ESTROMA

L'estroma [15] es pot definir com el marc de suport d'un òrgan (o glàndula o una altra estructura), generalment compost de teixit connectiu. És a dir, l'estroma és la trama (conformació d'un teixit integrat per cèl·lules i fibres) que sosté les cèl·lules. Per tant, està format per cèl·lules que li donen al teixit la seva força i forma. Moltes d'aquestes cèl·lules s'anomenen fibroblasts i també inclou vasos sanguinis que porten nutrients al teixit i canals limfàtics que eliminen l'excés de líquid i deixalles.

INFILTRACIÓ INFLAMATÒRIA

Els tumors atreuen un influx constant de cèl·lules inflamatòries que, quan es troben en el microambient tumoral, actuen com un mediador inflamatori que pot promoure el desenvolupament tumoral, induir la proliferació de cèl·lules iniciades i alterar l'estroma.

La resposta inflamatòria [16] dependrà de l'etapa del tumor (primera fase o metastàsica) i del millor o pitjor pronòstic. Aquesta resposta pot tenir efectes beneficiosos, generalment locals, inhibint el creixement tumoral, i efectes sistèmics que poden ser perjudicials. Per tant, el sistema immunològic pot tenir un important paper, tant a favor de la formació del tumor com en la inhibició de la mateixa en totes les seves etapes.

Així doncs, la resposta inflamatòria davant un esdeveniment tumoral pot augmentar i facilitar el creixement del tumor, ja que predisposa i promou el seu desenvolupament. Totes les cèl·lules que formen part del microambient tumoral participen en interaccions recíproques i contínuament van canviant les seves característiques, adquirint diferents papers adaptatius dins del propi microambient que poden tenir un efecte sobre el creixement del tumor, la seva progressió o la possible metàstasi.

NECROSI

La necrosi [17] és una mort prematura de les cèl·lules en un teixit o òrgan. En base a criteris morfològics i bioquímics, s'han definit tres classes de mort cel·lular: apoptosi, autofàgia i necrosi. Pel que fa a aquesta última, la mort d'una cèl·lula no és causada de forma natural, sinó per un mecanisme que involucra l'activació de certs receptors que resulta en la pèrdua de la integritat de la membrana cel·lular. Per tant, la necrosi es pot definir com la sèrie d'esdeveniments que condueixen a la ruptura de la membrana citoplasmàtica i la consegüent sortida de material intracel·lular a l'entorn extracel·lular.

S'ha entès la necrosi com una mort cel·lular no controlada que desencadena reaccions inflamatòries en els teixits circumdants i que pot afavorir la disseminació de patògens. És a dir, també s'associa als processos inflamatoris i, a més, en el càncer, afavoreix l'angiogènesi, el creixement tumoral i les metàstasis. Relacionat amb la necrosi, també trobem el concepte de *debris*, que fa referència a deixalles orgàniques que queden després de la mort d'una cèl·lula.

3.2.2. Proves diagnòstiques

Un dels aspectes clau és l'anàlisi per detectar el càncer de mama, inclosos la detecció, el diagnòstic i el control. Pel que fa la detecció, l'objectiu és detectar el càncer de mama aviat mitjançant proves com les mamografies. En canvi, les proves diagnòstiques, com les biòpsies, s'utilitzen per determinar si existeix càncer de mama

i, en cas afirmatiu, poder determinar la seva extensió. A més, també són útils per tenir més informació relativa al tractament. Finalment, les proves de control s'utilitzen per controlar l'eficàcia de les teràpies.

En el nostre cas, ens centrem en les biòpsies, que consisteixen en prendre una mostra de teixit de la zona afectada mitjançant una agulla o una incisió. Aquesta mostra és petita, i pot variar des d'un cilindre de la mida de la cànula de l'agulla fins a un cub de $1\text{cm} \times 1\text{cm} \times 1\text{cm}$. A continuació, es talla el teixit en diferents trossos i s'aplica a cadascun d'ells una tinció immunohistoquímica [18], que consisteix en un mètode per donar color a les cèl·lules, teixits o components microscòpics. D'aquesta manera, s'aconsegueix que ressaltin i es visualitzin millor, per exemple, a través d'un microscopi. En el cas del projecte DigiPatICS, les tincions disponibles són: *Cytokeratin 19* (CK19), Hematoxilina-Eosina (HE), marcadors de proliferació de KI-67 (KI-67), receptors de progesterona (PR o RP), receptors d'estrogen (ER o RE) i *Human Epidermal growth factor Receptor* (HER2), algunes de les quals les podem observar a la Figura 2 i a la Figura 3.

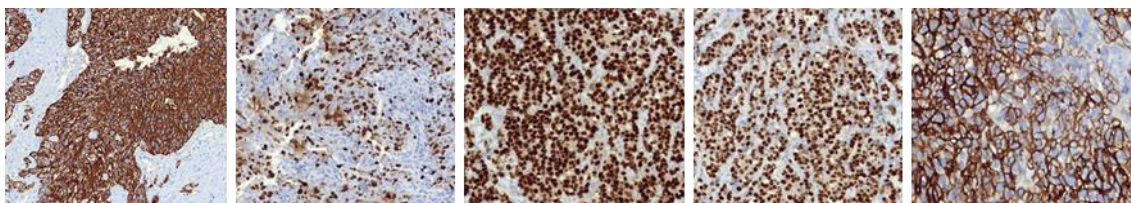


Figura 2. Exemples de les tincions CK19, KI-67, RE, RP i HER2.

Un cop s'han tenyit els trossos de teixit, s'escanegen per tal de generar diferents imatges digitals histològiques, amb el nom de *Whole Slide Images* (WSI). A continuació, els patòlegs analitzen el teixit per saber si conté cèl·lules tumorals o no. Aquesta cerca es fa inicialment amb la tinció HE, que és la que s'utilitza per detectar la presència de zones tumorals. Si es determina que en la mostra hi ha potencialment cèl·lules cancerígenes, s'analitzen les altres tincions per tal d'obtenir més informació rellevant en el diagnòstic. Per exemple, per marcar diferents aspectes relacionats amb els nuclis de les cèl·lules, s'observen les tincions KI-67, RP i RE. En canvi, per analitzar característiques de la membrana cel·lular, els patòlegs fan servir la tinció HER2. Aquesta informació s'obté només de les zones tumorals per tal d'obtenir una valoració més precisa basada en el número de cèl·lules positives i negatives (és a dir, les que sí han reaccionat a la tinció o no ho han fet), la intensitat dels nuclis tenyits i el percentatge de membrana tenyida i completa. S'ha de tenir en compte que no només hi ha les cèl·lules positives o negatives, sinó que també hi ha altres components explicats anteriorment com estroma, inflamació inflamatòria, necrosi, etc. Així doncs, a partir d'aquest elements, es pot decidir un tractament més òptim pel pacient.

Pel que fa aquest treball en concret, la tinció amb la qual es treballa és la Hematoxilina-Eosina (HE) [19] que, com ja s'ha esmentat anteriorment, és la que s'utilitza inicialment per detectar les zones tumorals en la mostra. Això és degut a que permet observar tant l'estructura cel·lular com la del teixit, ja que augmenta el contrast entre els diferents components per destacar algunes estructures específiques. Tal i com indica el seu nom, la tinció HE té dues components diferents, la Hematoxilina i l'Eosina. Mentre que l'Hematoxilina s'utilitza per mostrar detalls dels nuclis cel·lulars, l'Eosina s'utilitza principalment per diferenciar les components del citoplasma dels nuclis. Aquesta diferència entre Hematoxilina i Eosina es pot observar a la Figura 3, que mostra dos exemples de la tinció HE i la separació en les dues components.

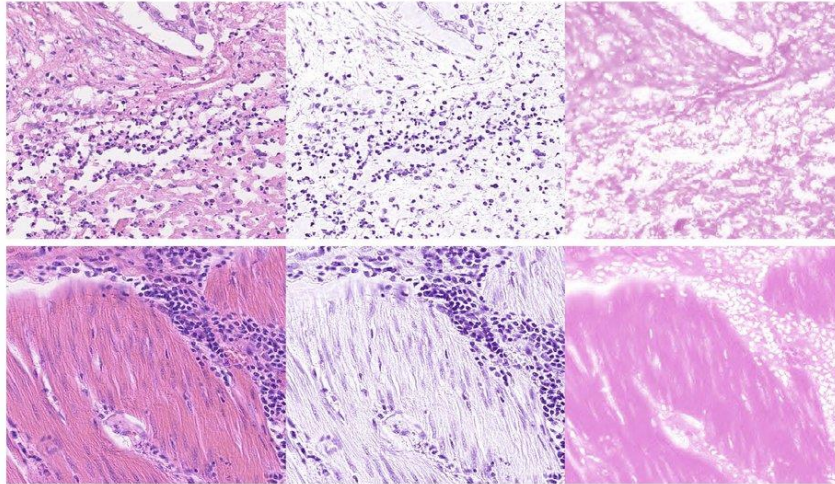


Figura 3. [20] Separació de la tinció HE. 1a columna: imatges HE, 2a columna: Hematoxilina, 3a columna: Eosina

3.2.3. Whole Slide Images (WSI)

Les *Whole Slide Images* (WSI) [21], [22], també conegudes com “microscòpia virtual”, són imatges digitals histològiques obtingudes mitjançant l’escaneig de portaobjectes per tal d’obtenir un arxiu digital d’alta resolució. Això s’aconsegueix capturant molts mosaics o tires d’imatges petites amb molt detall i després muntant-les per crear una imatge completa d’una secció histològica sencera. D’aquesta manera, els patòlegs poden examinar una mostra de biòpsia per tal de localitzar, analitzar i classificar malalties com el càncer.

L’objectiu de fer servir les WSI és emular la microscòpia òptica convencional d’una manera generada per ordinador. A més, d’aquesta manera s’obtenen arxius digitals d’alta resolució que es poden emmagatzemar, accedir, analitzar i compartir de manera eficient utilitzant tecnologies de gestió de portaobjectes. Aquest tipus d’imatges contenen seccions histològiques completes i permeten fer *zoom* per veure detalls a alta resolució com, per exemple, cèl·lules individuals. Per tant, les WSI permeten al patòleg examinar les seccions histològiques a diferents nivells de detall utilitzant diferent augments.

L’obtenció de les *Whole Slide Images* consta de dos processos. En el primer, s’utilitza maquinari especialitzat (escàner) per digitalitzar el portaobjectes i generar la imatge digital representativa (l’anomenada “diapositiva digital”). El segon procés empra programari especialitzat (és a dir, visor de diapositives virtual) per veure i/o analitzar aquests enormes arxius digitals.

Les WSI capturen molta informació ja que l’escaneig del portaobjectes s’obté amb molt detall. Per aquest motiu són molt grans, és a dir, tenen molts píxels. Per exemple, una *Whole Slide Image* típica pot contenir 100.000×100.000 píxels. En el cas d’aquest treball, s’utilitzen imatges de dues fonts: del *Breast Cancer Semantic Segmentation challenge* i de l’Institut Català de la Salut (ICS). En aquest últim cas, les imatges són de mida, aproximadament, 90.000×200.000 píxels.

Les *Whole Slide Images* poden estar en diferents extensions segons els formats utilitzats per guardar la informació obtinguda amb l’escaneig. Alguns d’aquests són: Aperio (.svs, .tif), MIRAX (.mrxs), Philips (.tiff), etc. En aquest projecte s’utilitzen dues

d'aquestes extensions: “.svs” i “.mrxs”. En el primer cas, es tracta d'un format en el qual hi ha un únic arxiu piramidal amb *tiles*, metadades i amb compressió no estàndard. En canvi, en el cas de la extensió “.mrxs”, aquesta consta de múltiples arxius amb metadades i índexs.

Per tal d'accedir a la informació d'una certa regió, és a dir, per veure una certa part de la imatge, es fa servir l'organització per *tiles*. La imatge es guarda en *tiles* que, a la seva vegada, es guarden per files, com es mostra a l'esquerra de la Figura 4. Tal i com es pot observar, els píxels es guarden a partir de la cantonada superior esquerra en *tiles* (representats pels quadrats de color lila). Per tant, per tal de llegir una certa regió de la imatge (píxels verd fosc), s'han de llegir tot els *tiles* que contenen algun dels píxels que conformen la regió (representats de color verd clar).

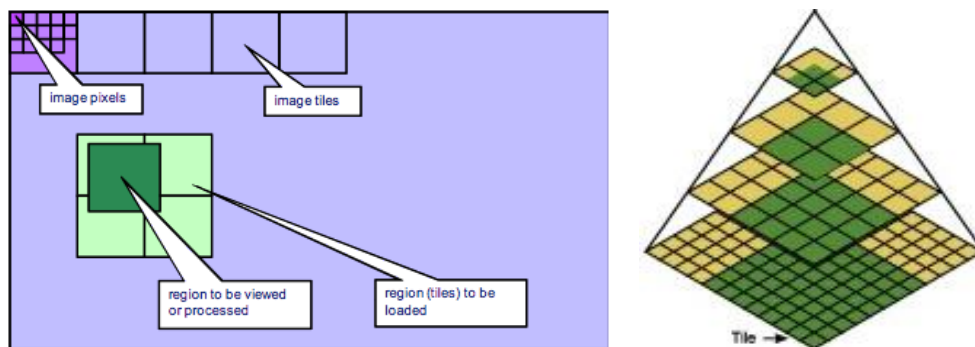


Figura 4. Whole Slide Image organitzada per tiles. Imatge piramidal amb els diferents nivells de resolució.

També s'han de tenir en compte els diferents nivells de resolució de les WSI. A partir de la imatge en màxima resolució, per tal d'obtenir la imatge a menys resolució el que es fa és concatenar 4 *tiles* del nivell anterior (resolució superior) i escalar amb un factor de 2, tal i com es mostra a la dreta de la Figura 4. Per exemple, una imatge de Vall d'Hebron de 90.000×200.000 píxels té una mida de 89×195 píxels en el nivell de mínima resolució, que correspon a un factor de *downsample* de 1.024. D'aquesta manera, es pot veure la *Whole Slide Image* amb més o menys detall i navegar còmodament sense haver de carregar tota la imatge sencera, fet que és impossible degut a la gran quantitat de píxels que tenen aquestes imatges.

3.3. Aspectes tecnològics

En aquesta subsecció introduïm la xarxa utilitzada, així com les mètriques d'avaluació i els *datasets* amb els que hem treballat. En el cas d'aquest projecte, es proposa fer segmentació dels *patches* extrets de les *Whole Slide Images*. Així doncs, s'explica l'arquitectura CNN utilitzada per fer segmentació d'imatges: la U-Net.

3.3.1. Arquitectura CNN

La U-Net [23]–[25] és una arquitectura de xarxa neuronal per fer segmentacions d'imatges de manera ràpida i precisa. La motivació d'aquest tipus de xarxa és que la segmentació semàntica no només necessita fer la discriminació a nivell de píxel, sinó que també s'han de projectar les característiques discriminatives a l'espai de píxels. És a dir, s'ha de reconstruir la imatge sencera a partir dels vectors de *features* extrets.

La seva arquitectura es pot pensar com un codificador seguit d'un descodificador. La primera part correspon a una xarxa per fer classificació. En canvi, la segona té com objectiu projectar les característiques discriminatives a l'espai de píxels per tenir una classificació densa. D'aquesta manera, les *features* apreses pel codificador s'utilitzen per tal de reconstruir la imatge en el descodificador. Entre aquestes dues parts també es pot considerar una tercera, la qual es pot pensar com un coll d'ampolla. Això es pot observar a la Figura 5, on es pot veure un diagrama de l'arquitectura de la U-Net amb les dues parts principals ben diferenciades.

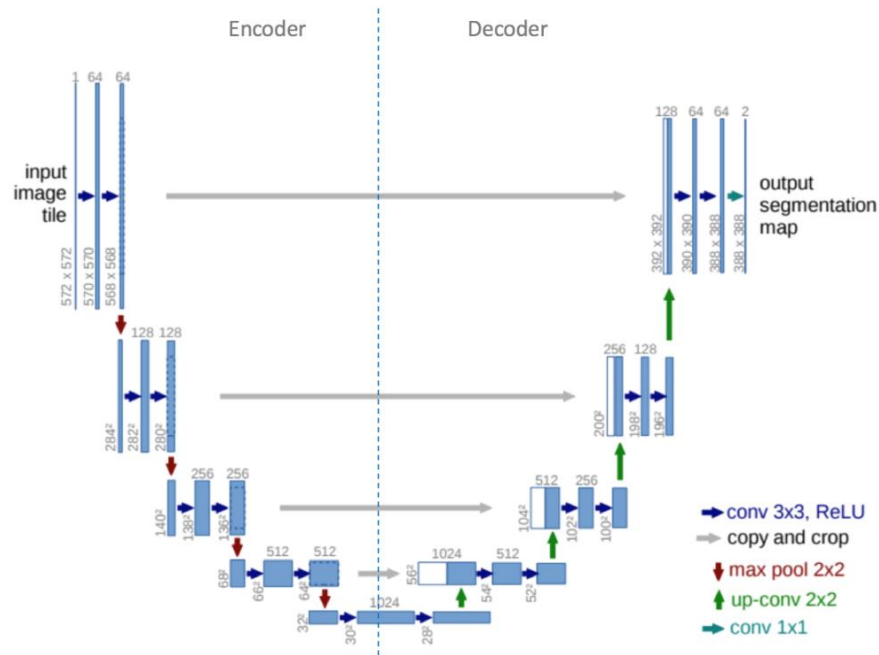


Figura 5. Estructura de la U-Net amb la part del codificador i del descodificador diferenciades.

El resultat de la U-Net representa la segmentació de la imatge original en C classes diferents. Concretament, la sortida obtinguda és de mida $C \times N \times M$, on N i M són les mides de la imatge passada com a entrada. Aquest *output* es passa per una capa d'activació (*softmax*, *sigmoid*, etc.) per tal d'obtenir un mapa de probabilitats. És a dir, es pot considerar que cadascun dels C canals conté, per cada píxel, la probabilitat de que sigui de la classe especificada. Per últim, s'obté el resultat final a partir de seleccionar la classe final com la més probable en cada cas. Així doncs, s'obté la màscara amb la segmentació on, en cada posició, s'indica la classe més probable.

Per tant, la funció d'activació s'aplica com a última capa de la xarxa neuronal per tal de normalitzar la sortida del model i obtenir una distribució de probabilitat sobre les classes de sortida. Una de les més utilitzades és la Softmax [26], que calcula un vector de la probabilitat de cadascuna de les classes possibles a partir de la Eq.(1). Així doncs, les components d'aquest vector sumen 1.

$$S(y) = \frac{e^{y_i}}{\sum e^{y_j}} \quad (1)$$

A l'hora d'entrenar la xarxa neuronal, s'han de tenir molt presents diversos aspectes que són fonamentals. El primer d'ells és l'*encoder* a fer servir per tal d'extreure les *features* de la imatge, el segon correspon a la *loss function* o funció de pèrdues i, finalment, les mètriques utilitzades per avaluar el model.

ENCODERS

L'encoder és un model classificador que s'utilitza en l'arquitectura de la U-Net per extreure les *features*. Així doncs, aquest pot ser una ResNet, VGG, DenseNet, EfficientNet, etc. És usual que el codificador sigui una xarxa pre-entrenada, és a dir, que els pesos de les diferents capes estiguin inicialitzats de manera no aleatòria, sinó amb els valors resultants del model òptim d'una altra tasca.

Com s'ha esmentat anteriorment, hi ha diferents tipus d'encoders segons les característiques de cadascun. En aquest projecte, s'han utilitzat només alguns dels possibles models classificadors, i que són els que s'expliquen a continuació.

1. ResNeXt

La ResNeXt [27], [28] es tracta d'un classificador que consisteix en apilar blocs residuals, que agreguen un conjunt de transformacions amb la mateixa topologia. És a dir, afegeix una nova dimensió respecte la ResNet, la cardinalitat C , que fa referència a la mida del conjunt de transformacions a ser agregades. Per tant, en comptes d'augmentar la profunditat, el que es fa en aquesta xarxa és expandir en una nova dimensió anomenada cardinalitat. Per aquest motiu és una arquitectura multi-branca, tal i com es pot veure a la Figura 6.

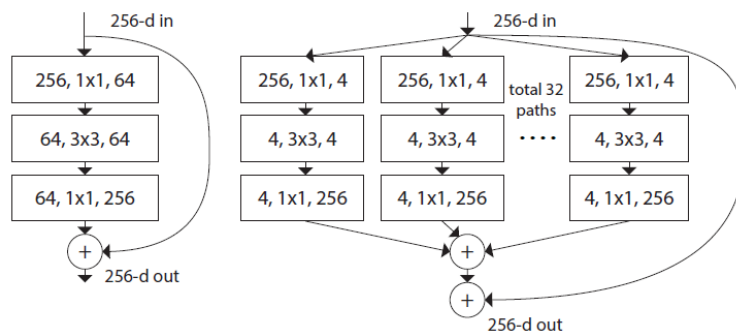


Figura 6. Esquerra: Bloc de la ResNet. Dreta: Bloc de la ResNext amb cardinalitat=32 amb la mateixa complexitat. Cada capa es representa com (# canals d'entrada, mida del filtre, # canals de sortida).

2. DenseNet

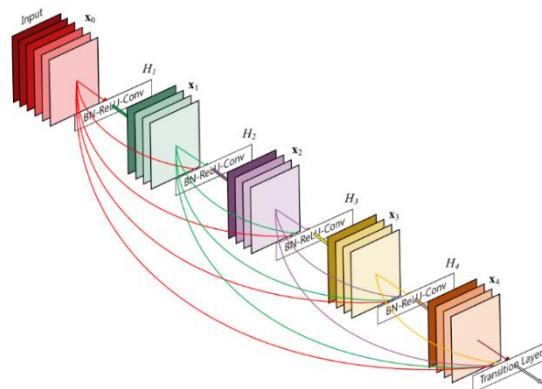


Figura 7. Bloc dens de 5 capes on cada capa rep com entrada tots els feature-maps de les capes anteriors.

La *Dense Convolutional Network* (DenseNet) [29] connecta cada capa a cadascuna de la resta de capes d'una manera *feed-forward*. Per tant, tal i com es pot veure a la Figura 7, per cada capa, s'utilitzen com entrada tots els *feature maps* de les capes

prèvies, i el *feature map* de sortida formarà part de l'entrada de totes les capes posteriors. La idea és assegurar el màxim flux d'informació entre les capes de la xarxa, ja que es connecten totes les capes de la xarxa entre elles.

3. Xception

Aquesta xarxa està basada en la Inception, per aquest motiu, s'anomena Extrem Inception (Xception) [30], [31]. La seva arquitectura es pot definir com una pila de capes convolucional separables en termes de profunditat amb connexions residuals, tal i com es pot veure a la Figura 8. Per tant, es tracta d'una capa de convolució de mida 1×1 seguida d'una capa convolucional per cadascun dels canals per separat.

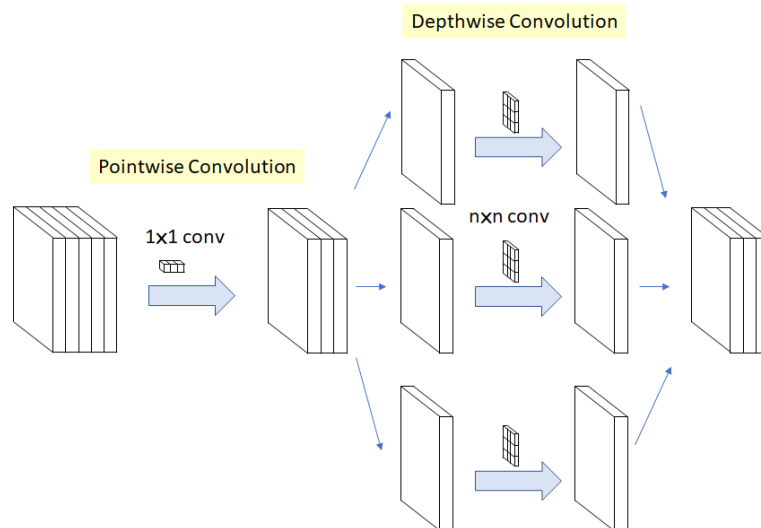


Figura 8. Convolucions separables en profunditat (canals).

4. EfficientNet

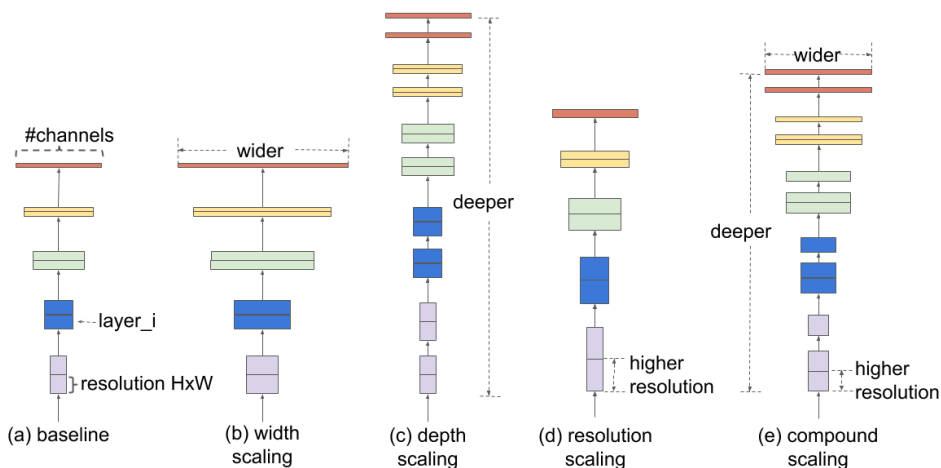


Figura 9. Diferents mètodes d'escalat. (a) fa referència al model base; (b)-(d) són escalats convencionals que només augmenten una dimensió; (e) mostra l'escalat compost proposat per tal d'escalar uniformement les tres dimensions.

Les xarxes neuronals convolucional es desenvolupen normalment a un cost de recursos fix i després s'amplien per aconseguir una millor precisió quan hi ha més recursos disponibles. L'EfficientNet [32], [33] es tracta d'un mètode que escala uniformement l'amplada, la profunditat i la resolució de la xarxa amb un conjunt de coeficients. Les diferents possibilitats d'escalat es poden observar a la Figura 9, on es mostren els efectes dels escalats de l'amplada (número de neurones en una capa), la

profunditat (número de capes de la xarxa) i la resolució (l'alçada i l'amplada de la imatge d'entrada).

Aquests models no només proporcionen una millor precisió, sinó que també milloren l'eficiència dels models al reduir els paràmetres i el número de FLOPS (*Floating Point Operations Per Second*) en comparació amb els models anteriors.

5. VGG

En el cas de la VGG [34], [35], el canvi respecte altres models es centra en l'efecte de la profunditat de les xarxes convolucionals. Per aquest motiu, el que es proposa és augmentar constantment la profunditat de la xarxa afegint més capes convolucionals, la qual cosa és factible a causa de l'ús de filtres de convolució molt petits en totes les capes, concretament de mida 3×3 .

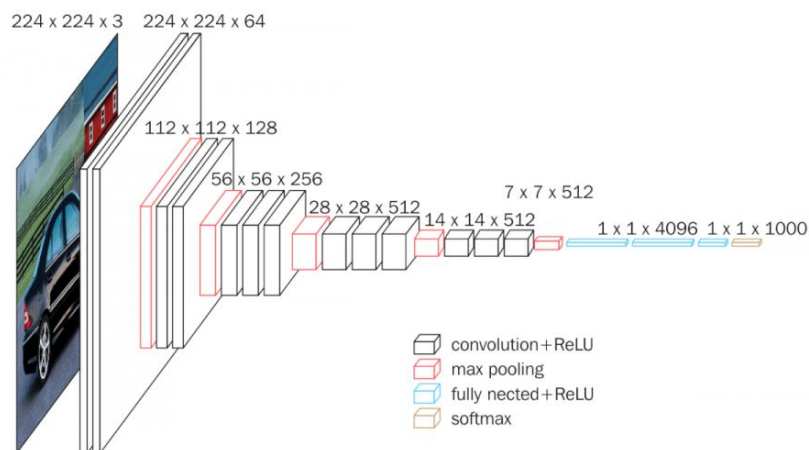


Figura 10. Arquitectura de la xarxa VGG16.

LOSS FUNCTION

La *loss function* o funció de pèrdua mesura la qualitat d'un conjunt de paràmetres comparant les prediccions respecte el valor esperat, és a dir, el del *ground truth*. Per tant, avalua el model basant-se en si coincideixen les sortides de la xarxa amb els valors correctes de les anotacions. A més, s'utilitza per guiar el procés d'entrenament per trobar el conjunt de paràmetres que minimitza el valor de la funció de pèrdua. Dit d'una altra manera, l'objectiu de l'entrenament és minimitzar la *loss function*.

Hi ha diferents tipus de *loss functions* segons l'objectiu: classificació o regressió. En el nostre cas, ens centrem en les del primer tipus, ja que la xarxa prediu variables categòriques corresponents a les diferents classes. En realitat, però, la solució proposada és fer segmentació d'imatges i, per tant, es tracta d'un model multi-classe amb C classes. Per aquest motiu, com s'ha esmentat a l'apartat anterior, la sortida del model conté, per cada píxel de la imatge d'entrada, C prediccions, una per cadascuna de les categories possibles. Així doncs, les possibles funcions de pèrdua utilitzades en aquest projecte han de poder ser utilitzades en aquestes circumstàncies.

A continuació, s'expliquen les diferents funcions de pèrdues amb les que s'ha treballat en aquest projecte, especificant el càlcul del valor a partir de les imatges i de la matriu de confusió. Per tal de representar la predicció i la màscara s'utilitzen les lletres A i B , mentre que per representar els elements de la matriu de confusió es fan servir les abreviacions TP (*True Positive*), FP (*False Positive*), FN (*False Negative*) i TN (*True Negative*).

1. Jaccard Loss

En aquest cas, el que s'utilitza en realitat per jutjar el model és una mètrica, l'índex de Jaccard o la *Intersection Over Union* (IoU). Per tant, es deriva la funció de pèrdues d'aquest valor fent $1 - f(x)$, on $f(x)$ és la mètrica corresponent, la IoU.

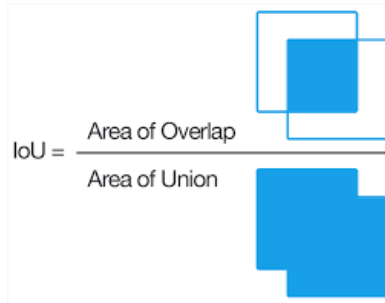


Figura 11. Intersection Over Union o índex de Jaccard.

L'índex de Jaccard o la IoU calcula, tal i com es pot veure a la Figura 11, el rati entre l'àrea de superposició de les prediccions i el *ground truth*, i l'àrea d'unió, és a dir, l'àrea abastada per les dues. Per tant, per obtenir la *Jaccard loss*, s'ha de calcular el valor de $1 - J(A, B)$, obtingut a partir de la Eq.(2)

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{TP + FP + FN} \quad (2)$$

2. Dice Loss

Com abans, aquesta funció de pèrdua també està basada en el càlcul d'una mètrica i, per tant, el seu valor es deriva d'un altre, fent $1 - DSC(A, B)$. En aquest cas, en canvi, el que s'utilitza és el coeficient Dice o Dice-Sørensen. El valor es calcula mitjançant la Eq.(3), és a dir, a partir de l'àrea de superposició de les prediccions i el *ground truth*, i la quantitat total de píxels en les dues imatges, tal i com representa la Figura 12.

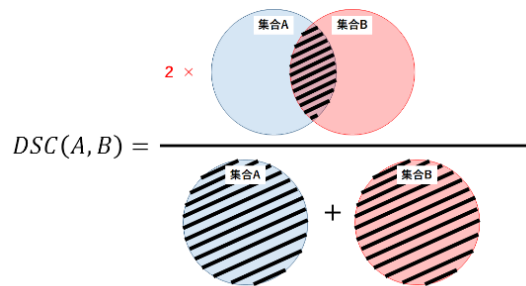


Figura 12. Coeficient Dice.

$$DSC(A, B) = \frac{2 |A \cap B|}{|A| + |B|} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (3)$$

Respecte la diferència entre les dues mètriques [36], [37], la *Intersection Over Union* tendeix a penalitzar instàncies mal classificades més que el coeficient de Dice. Per aquest motiu, es pot interpretar que aquest últim tendeix a mesurar de manera més propera el *performance* mitjà, mentre que la IoU és una mesura més relacionada al pitjor cas, és a dir, al valor mínim de les mètriques.

3. Cross Entropy Loss

La *Cross Entropy Loss* [38], [39] mesura la classificació d'un model tenint en compte que la sortida d'aquest correspon a la probabilitat de que una mostra pertanyi a una certa classe, per tant, a valors entre 0 i 1. Per aquest motiu es diu que aquesta funció de pèrdua té una interpretació probabilística, ja que es pot pensar com que mesura la diferència entre dues distribucions de probabilitat diferents: les prediccions i el *ground truth*.

$$CE = - \sum_{i=1}^n y_i \log(p_i) = -\log(p_t) \quad (4)$$

Aquesta *loss* es calcula de tal manera que es penalitza predir amb poca probabilitat la classe correcta, és a dir, es castiga que el model tingui poca confiança en la predicció d'aquesta categoria. Per obtenir el valor s'utilitza l'expressió de la Eq.(4), on y_i representa el *label* de cada classe (*one-hot encoding*) i p_i és la probabilitat obtinguda després d'aplicar la funció d'activació. Com es pot veure, aquesta expressió es pot simplificar utilitzant p_t , que correspon a la probabilitat de la classe correcta. Això és degut a que la única probabilitat que es té en compte realment en el càlcul de la *loss* és la que ha obtingut el model per la classe real.

$$WCE = - \sum_{i=1}^n w_i y_i \log(p_i) = -w_j \log(p_t) \quad (5)$$

Un fet a tenir en compte és si el *dataset* està balancejat o no, és a dir, si té la mateixa quantitat de dades en cadascuna de les classes o no. En aquest últim cas, això fa que el model aprengui més sobre les classes més freqüents. Per tant, en el cas de tenir un *dataset* desbalancejat, s'utilitza la *Weighted Cross Entropy Loss*, representada per la Eq.(5) que calcula la funció de pèrdues utilitzant uns pesos assignats a cadascuna de les classes de manera inversament proporcional a la seva freqüència. És a dir, que el *label* més habitual és el que té un pes associat més petit mentre que la classe menys freqüent té un pes màxim. D'aquesta manera es té en compte la quantitat de vegades que apareix cadascuna de les classes en les dades i es fa que el model es fixi més en les classes poc representades que en les més freqüents.

MÈTRQUES

A continuació, s'expliquen les mètriques que s'han utilitzat per comparar els diferents models entrenats en aquest projecte. Concretament, es tracta de quatre mètriques diferents que es calculen a partir de la matriu de confusió: el *recall*, la precisió, el valor de *F-Score* i l'*accuracy* [40]–[42].

Primer de tot, les matrius de confusió [43] són matrius de mida $C \times C$ que s'utilitzen per avaluar el model, on C és el número de classes. El que fa es comparar els valors anotats (els del *ground truth*) amb les prediccions del model de manera que és possible tenir una visió general de la *performance* del model i dels errors més comuns. Per exemple, si les columnes d'aquesta matriu representen els valors predits i les files corresponen a les anotacions, cadascun dels valors a_{ij} de la matriu indica la quantitat d'observacions anotades amb el *label* i que han estat classificades com pertanyents a la classe j .

A partir d'aquesta matriu, s'obtenen diferents tipus d'elements de la matriu de confusió, abreuiats com *TP* (*True Positive*), *FP* (*False Positive*), *FN* (*False Negative*) i *TN* (*True Negative*). A partir d'aquests valors es calculen la mètriques esmentades anteriorment, tal i com es pot observar en cadascuna de les equacions corresponents. S'ha de tenir en compte que, en el context multi-classe, aquestes mètriques es calculen a partir de comparar els valors d'aquesta i els de tota la resta de classes.

La primera mètrica és el *recall*, que informa sobre la sensitivitat del model, és a dir, la quantitat d'identificacions per part del model. Per obtenir aquest valor, es calcula quants dels elements del *ground truth* han estat predits pel nostre model, i identificats correctament. Per tant, es correspon al rati entre les prediccions correctes i totes les observacions anotades de la classe, tal i com es pot veure a la Eq.(6). Per tant, és una mètrica important sobretot quan hi ha un gran cost associat a falsos negatius.

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

La precisió s'utilitza per mesurar la qualitat del model, ja que indica com de precís és el model respecte les seves prediccions. És a dir, de les prediccions fetes pel model, avalua quantes d'elles són també positives en el *ground truth*. Com es pot veure a la Eq.(7), es calcula a partir del rati entre les prediccions positives correctes respecte totes les prediccions del model. Per tant, aquesta mesura és recomanable quan el cost dels falsos positius és alt.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

El valor del F1-Score s'utilitza per tal de balancejar la precisió i el *recall*. D'aquesta manera, es pot comparar el rendiment dels dos aspectes a la vegada. En el cas de F1, les dues mètriques tenen el mateix pes, per tant, es calcula a partir de la mitjana harmònica, mitjançant la Eq.(8)

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Per últim, l'*accuracy* mesura el percentatge de casos en els que el model ha encertat i es calcula a partir de la Eq.(9). Aquesta mètrica, però, no funciona bé en els casos en els que les classes són desbalancejades, ja que pot passar que un model dolent tingui un valor d'*accuracy* molt alt. En aquest cas, s'utilitza la *balanced accuracy*, que es calcula com el promig dels valors del *recall* per classe.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

3.3.2. Breast Cancer Semantic Segmentation (BCSS) dataset

El *Breast Cancer Semantic Segmentation* (BCSS) *dataset* [44]–[46] forma part d'un *challenge* que conté anotacions de segmentació de regions de teixit amb càncer de mama. Les imatges d'aquest *dataset* estan anotades de forma bastant precisa i les classes anotades en aquesta base de dades són molt completes. Per tant, es pot considerar que les imatges disponibles permeten entrenar models de *machine learning* acurats per tasques de segmentació.

A més, com s'ha esmentat anteriorment, l'anotació d'imatges histològiques és una tasca molt tediosa en la que es necessita experiència per tal de delinear correctament les diferents estructures presents en el teixit. Per aquest motiu, és molt difícil trobar un conjunt d'imatges anotades de manera tan acurada i definint cadascuna de les regions de les diferents classes.

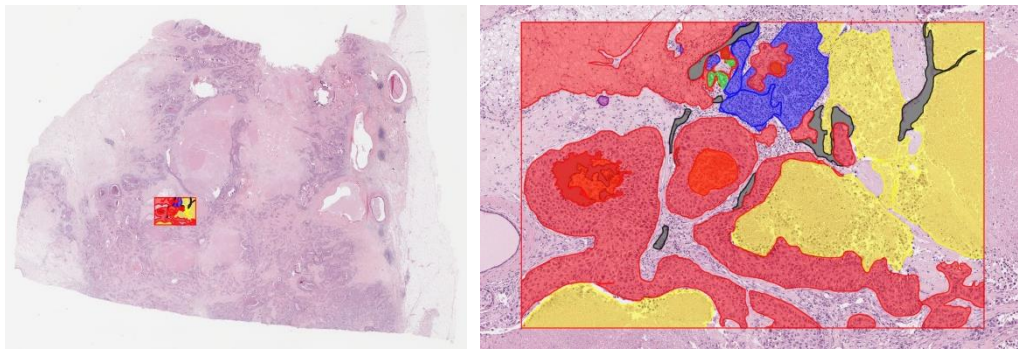


Figura 13. Exemple d'una *Whole Slide Image* i la seva ROI anotada del BCSS *dataset*.

Aquest base de dades consta de 151 *Whole Slide Images* (WSI) tenyides amb Hematoxilina i Eosina (HE) que corresponen a diferents casos de càncer de mama confirmats. En realitat, les anotacions són de regions d'interès (ROIs) de cadascuna de les WSI i que han estat seleccionades de manera que siguin el màxim de representatives de cadascuna de les imatges histològiques, i de manera que tinguin una densitat de tumor alta, si és possible. A la Figura 13 es pot observar una de les WSI de la base de dades conjuntament amb la ROI seleccionada i les anotacions corresponents de cadascun dels elements pertanyents a aquesta àrea. Per tant, en realitat el *dataset* consta de 151 imatges corresponents a una ROI de cadascuna de les WSI seleccionada de manera que la classe corresponent al tumor estigui força representada.

Les imatges que formen part d'aquest *dataset* han sigut extretes de l'Atlas del Genoma del Càncer (*The Cancer Genome Atlas*, TCGA) i totes elles són del tipus triple negatiu, és a dir, que han donat negatiu en les tincions RE, RP i HER2. Aquest últim tret no té efecte en les característiques pròpies de les diferents classes i, per tant, no suposa cap problema per a la utilització del nostre model en qualsevol tipus de tumor.

Les imatges han estat generades seguint diversos passos representats a la Figura 14 en els quals han col·laborat patòlegs, residents i estudiants de medicina. Primer de tot, s'ha seleccionat una ROI a anotar per cadascuna de les *Whole Slide Images*. A continuació, s'han reclutat 25 participants als quals se'ls ha fet diverses sessions d'entrenament per tal d'explicar les classes a anotar i la utilització de l'eina a fer servir per l'anotació, *Digital Slide Archive* [47]. El següent pas, és l'assignació de les diferents

imatges a cadascun dels participants, adjudicant 5 o 6 imatges a cadascun d'ells. Aquest repartiment es fa de manera que les *WSI* amb més dificultat s'assignen a patòlegs més experimentats. A continuació, durant el procés d'anotació amb *Digital Slide Archive*, els participants poden rebre *feedback* sobre les anotacions i fer preguntes en cas de dubte. Després d'aquest pas, s'utilitzen unes ROIs comunes per tots els participants per tal de comparar entre anotadors. Per últim, es mesura mitjançant el coeficient de Dice la discordança entre les anotacions dels diferents participants.

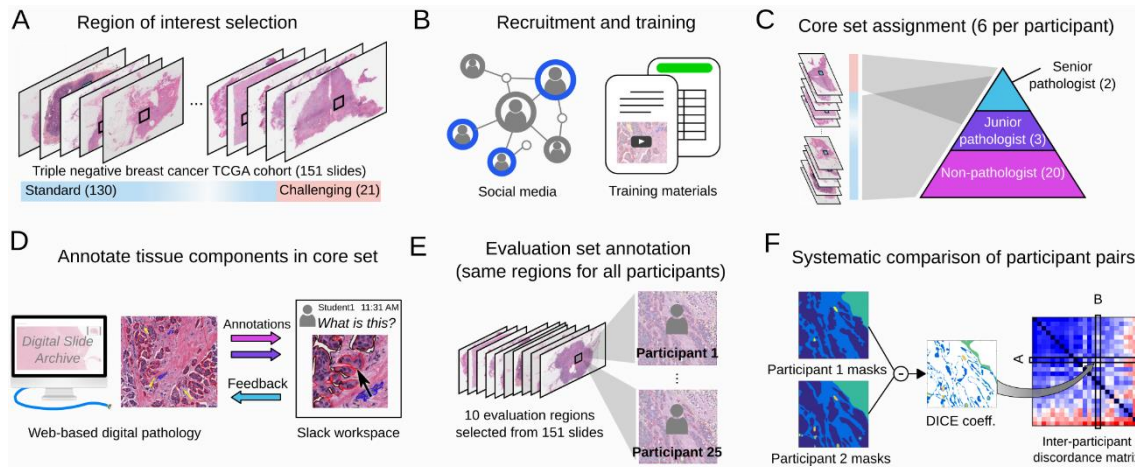


Figura 14. Procés d'obtenció de les anotacions de les imatges.

3.3.3. Dades de l'ICS

En el cas de l'ICS (Institut Català de la Salut), per a la realització d'aquest projecte s'ha disposat de les *Whole Slide Images* de 20 pacients diferents, cadascun d'ells amb diferents tincions. En el cas de l'Hematoxilina-Eosina (HE), hi ha 14 *WSI* disponibles. Concretament, totes les imatges histològiques disponibles són de l'Hospital Vall d'Hebron.

Tot i poder accedir a aquestes imatges, no s'han pogut aconseguir anotacions de diferents regions d'aquestes *WSI* per tal d'entrenar la xarxa. És per aquest motiu que, per entrenar el model, s'ha fet servir el *dataset* explicat anteriorment, el *Breast Cancer Semantic Segmentation* (BCSS), ja que aquest sí disposa d'anotacions precises de diferents classes, amb l'estratègia d'emprar el model entrenat amb BCSS per poder fer inferència sobre imatges de l'ICS.

Cal destacar que, com s'ha explicat anteriorment, un dels problemes a tenir en compte en aquest projecte és la intervariabilitat de les *WSI*, ja que diferents procediments, productes químics i eines utilitzades a l'hora de tractar les mostres poden produir diferents tipus d'imatges. Tal i com es pot veure a la Figura 15 i a la Figura 16, les mostres del BCSS i les de l'ICS (concretament les de Vall d'Hebron) no s'assemblen gaire i, per tant, això pot suposar un problema a l'hora de fer inferència sobre les imatges de l'Institut Català de la Salut.

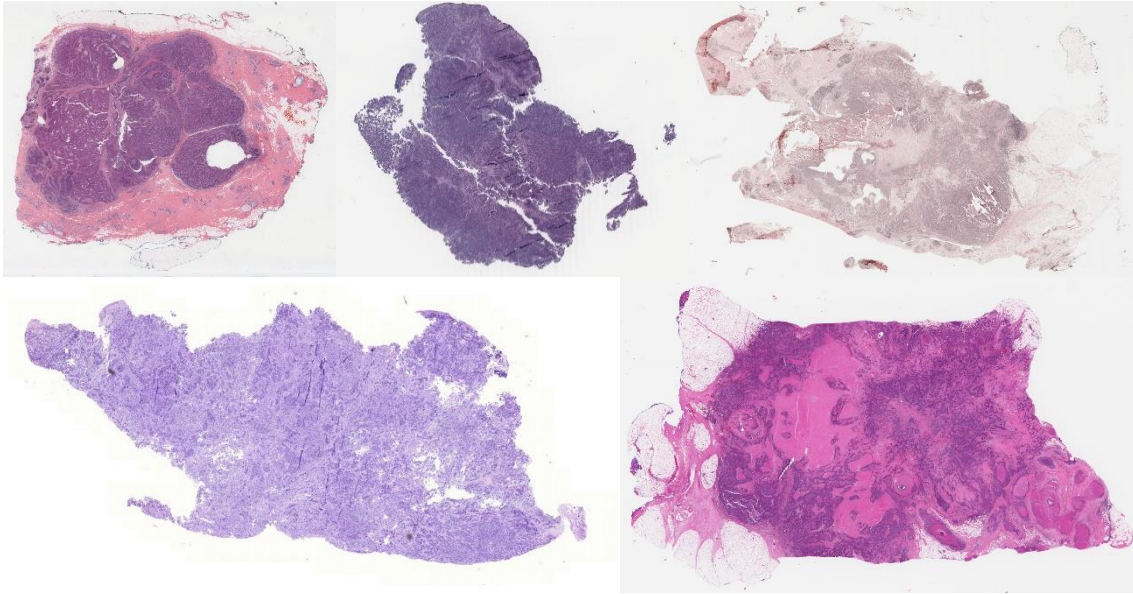


Figura 15. Exemples de WSI del Breast Cancer Semantic Segmentation dataset.



Figura 16. Exemples de WSI de la base de dades de l'Institut Català de la Salut.

4. Solució proposada

Tradicionalment s'han considerat dos enfocaments diferents amb xarxes neuronals per detectar tumors de càncer de mama en imatges histològiques: fer classificació de *patches* i fer segmentació semàntica. En els dos casos, la idea és dividir la imatge original, la *Whole Slide Image*, en *patches* per tal de tenir imatges de diferents regions de teixit d'una mida fixa amb els quals es pot treballar sense problemes de memòria.

La classificació, per tant, consisteix en decidir la classe de cadascun d'aquests *patches* per tal d'obtenir un "mapa" final de la WSI sencera i que els patòlegs puguin veure en quines regions és més probable que hi hagi cèl·lules tumorals. En canvi, en la segmentació, per cadascuna d'aquestes regions, el resultat seria una classificació a nivell de píxel per tal d'obtenir una detecció precisa de la zona tumoral, si n'hi ha. En la Figura 17 es poden veure dos exemples diferents de resultats tant en la classificació de *patches* com en la segmentació de regions, on els colors en cada cas representen classes diferents.

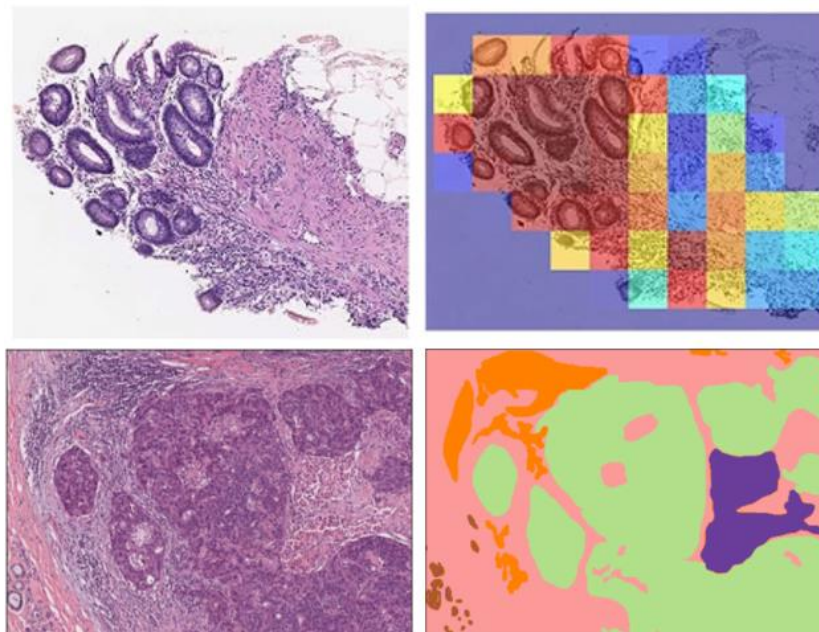


Figura 17. Classificació de *patches* i segmentació d'una regió.

En el cas d'aquest treball, es proposa seguir l'enfocament de segmentació de les diferents regions extretes de les *Whole Slide Images*. Això és possible ja que s'ha pogut disposar d'una base de dades anotades de mostres de càncer de mama per fer segmentació de WSI tenyides amb Hematoxilina-Eosina, anomenada BCSS. A més, s'ha considerat que és més útil pels patòlegs obtenir una segmentació el més acurada possible de cadascuna de les regions amb les possibles classes en comptes d'una classificació individual de cadascun dels *patches*.























Respecte les eines utilitzades en la realització d'aquest treball, s'ha fet servir Python con el llenguatge de programació principal en les diferents tasques realitzades. Per tal d'entrenar la xarxa neuronal, s'ha fet servir la llibreria *Segmentation Models Pytorch* [48], la qual proporciona diferents models per fer segmentació binària i multi-classe. Per últim, també és important destacar que els entorns utilitzats per executar els codis

en aquest projecte han sigut Google Colab i el servidor de GPI (Grup de Processament d'Imatge).

4.1. BCSS dataset

Com s'ha explicat a l'apartat 3.3.2, aquest *dataset* consta de 151 imatges amb les màscares de la classe corresponent de cada píxel. Un aspecte important i que ens aporta més flexibilitat a l'hora d'entrenar el model és que, al descarregar les imatges i les màscares, es pot especificar la magnificació a la qual es volen obtenir. Per tant, es pot disposar de les imatges amb diferents nivells de resolució, és a dir, amb més o menys detall.

Taula 1. Labels i codis abans i després de l'agrupació en les classes finals.

Labels – GT	Code - GT	Code - 5 classes	Labels - 5 classes	Color
outside_roi	0	0	outside_roi	
tumor	1	1	tumor	
stroma	2	2	stroma	
lymphocytic_infiltrate	3	3	inflammatory infiltrates	
necrosis_or_debris	4	4	necrosis	
glandular_secretions	5	5	other	
blood	6	5	other	
exclude	7	5	other	
metaplasia_NOS	8	5	other	
fat	9	5	other	
plasma_cells	10	3	inflammatory infiltrates	
other_immune_infiltrate	11	3	inflammatory infiltrates	
mucoid_material	12	5	other	
normal_acinus_or_duct	13	5	other	
lymphatics	14	5	other	
undetermined	15	5	other	
nerve	16	5	other	
skin_adnexa	17	5	other	
blood_vessel	18	5	other	
angi Invasion	19	1	tumor	
dcis	20	1	tumor	
other	21	5	other	

Les anotacions originals consten de 21 classes diferents més una addicional per indicar que els píxels no pertanyen a la ROI i que, per tant, no estan anotats. En el *paper* del *challenge* [46] els autors esmenten que aquestes classes es poden reduir a 5 per tal d'entrenar el model de segmentació. És important dir que aquesta agrupació proposada s'ha mostrat als metges del projecte DigiPatICS per tal de conèixer la seva opinió quan es faci inferència més endavant amb el *dataset* de l'ICS i, d'aquesta manera, poder fer modificacions a l'hora d'agrupar. L'únic canvi que se'ns va proposar va ser fer una classe pròpia per "dcis" (carcinomes ductals in situ), però no ha sigut possible ja que només hi ha una imatge anotada amb aquest *label*. Per tant, el conjunt de dades amb el que s'ha treballat consta de les 5 classes proposades més

l'addicional que s'utilitza per indicar que els píxels no han estat anotats. En la Taula 1 es poden veure els *labels* amb els seus codis corresponents en les anotacions originals, i també després de fer l'agrupació en les classes especificades pels autors del *Breast Cancer Semantic Segmentation challenge*. A la Figura 18 es mostren dos exemples d'imatges i les seves màscares després de fer l'agrupació en les classes finals i que, per tant, són exemples del *dataset* que s'ha utilitzat finalment per entrenar el model de segmentació.

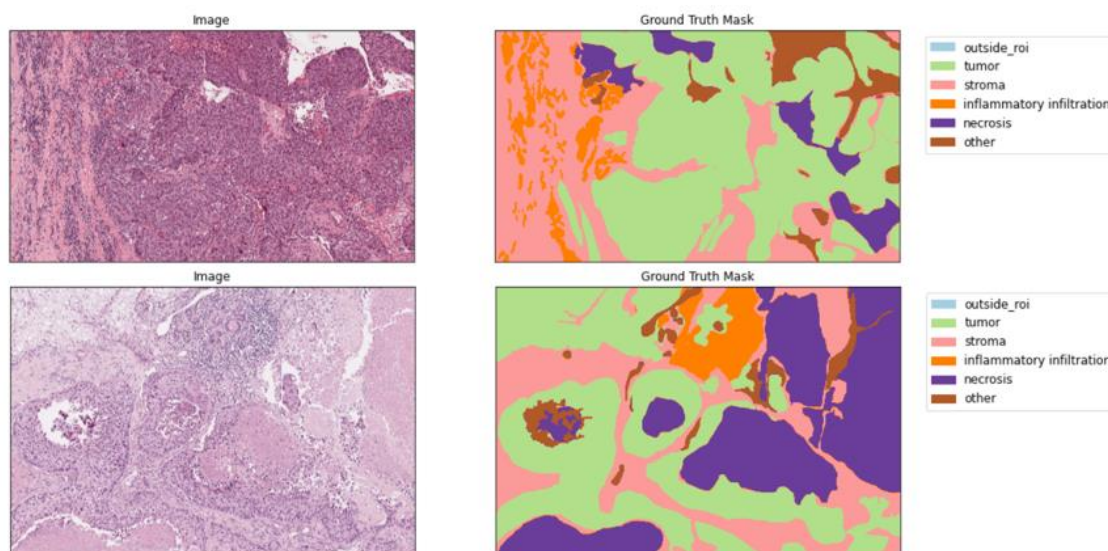


Figura 18. Exemples de les imatges i màscares que s'usaran per entrenar el model de segmentació.

Per tal d'entrenar la xarxa, s'ha construït el *dataset* final separant aquestes imatges en tres conjunts: d'entrenament, de validació i de test. Per fer aquesta divisió, s'ha seguit la recomanació del *challenge* [44], que fa una separació del 70% d'imatges d'entrenament i un 30% de test, especificant quines imatges s'han utilitzat en cada conjunt, no només la proporció de la divisió. En el cas d'aquest projecte, de les imatges que conformen originalment el conjunt de test, s'han dividit de manera que s'ha obtingut un conjunt de validació i un altre de test. Així doncs, el percentatge d'imatges en cadascun dels conjunts finals d'entrenament, validació i test és del 70%, 15% i 15%, respectivament.

Aquestes imatges, corresponents a les diferents ROIs, tenen una mida diferent en cada cas. Per exemple, en el cas de magnificació màxima (40), algunes imatges són de mida 6.311×4.845 , 11.040×10.220 , 5.086×3.628 , etc. Per tal de què el model pugui disposar de més dades, en el cas de les imatges d'entrenament i de validació, el que s'ha fet és obtenir diversos *patches* de cadascuna d'elles. Per tant, cadascuna d'aquestes imatges s'ha retallat de manera que la mida resultant és, per cada dimensió, el mínim múltiple de 1.000 més proper. A continuació, la imatge resultant s'ha dividit de manera que s'obtenen diferents *patches* de mida 1.000×1.000 . Per exemple, en els cassos anteriors, les imatges resultants del *crop* tenen mida 6.000×4.000 , 11.000×10.000 i 5.000×3.000 , i d'elles s'obtenen 24, 110 i 15 *patches* de mida 1.000×1.000 píxels, respectivament.

Per tal d'evitar la pèrdua d'informació útil, el *crop* s'ha fet de manera que es retalli l'àrea amb màscara si aquesta no està buida i, en cas contrari, fer-ho aleatòriament. En el nostre treball, es considera que la màscara està buida si l'anotació pertany a la

classe “outside_roi”, que indica que aquests píxels no han estat anotats. Per tant, s’intenta agafar el màxim de regió anotada si és possible.

També s’ha considerat mantenir tota la informació fent *padding* de la imatge per tal de que la mida resultant fos el múltiple de 1.000 límit superior de cadascuna de les dimensions. D’aquesta manera, els valors dels píxels afegits a les vores de la imatges s’obtenien mitjançant *mirroring*. Aquest fet, però, introduïa patrons artificials a les imatges que no eren naturals i, finalment, s’ha descartat aquesta opció de *padding*.

Per tant, amb tot això s’ha generat el *dataset* final format per tres conjunts diferents: entrenament, validació i test, de manera que els dos primers estan formats per *patches* de mida 1.000 × 1.000 obtinguts de les imatges originals corresponents, i el conjunt de test està format per les imatges originals corresponents.

A l’hora d’entrenar la xarxa, però, no s’han utilitzat les imatges directament, sinó que s’han aplicat diversos mètodes de *data augmentation* [49] per tal de crear noves mostres de *training* modificant les existents. D’aquesta manera s’ajuda a evitar el sobreentrenament i es millora el rendiment del model. En el cas de les dades d’entrenament, les transformacions que s’han aplicat han sigut les següents:

- Crop: en aquest cas, es retallen trossos d’una certa mida intentant evitar agafar la part de la màscara anotada com “outside_roi”, si existeix; si no, s’agafa un tros aleatòriament.
- Flip: es tracta de donar la volta a la imatge horitzontalment, verticalment o de les dues maneres a la vegada. En aquest cas, la transformació no s’aplica sempre, sinó que només ho fa amb un 50% de probabilitats en cada cas.
- Rotacions de 90°: la imatge es pot rotar segons un angle múltiple de 90°, és a dir, 0, 90, 180 i 270° (considerant els angles mòdul 360°). Com en la transformació anterior, aquesta només s’aplica en un 50% dels casos.

En canvi, en el cas de les dades de validació, només s’aplica un *padding* per fer que la mida de les imatges sigui múltiple de 32. Això es fa perquè no hi hagi problemes a la xarxa, a la U-Net, a l’hora d’augmentar la mida dels tensors en cada pas del *decoder*.

Un altre factor a tenir en compte al passar les imatges a la xarxa per entrenar el model, és que s’ha aplicat una normalització a les dades. Concretament, s’ha utilitzat tant la mitjana com la desviació estàndard calculada a partir de les imatges d’entrenament originals, és a dir, les ROIs anotades del *dataset* original. El valors de la mitjana i la desviació estàndard que s’han utilitzat per normalitzar les dades són [0.7197 0.5453 0.6843] i [0.1973 0.2321 0.1732], respectivament.

4.2. Arquitectura CNN

Com s’ha explicat anteriorment, la U-Net és una arquitectura per fer segmentació d’imatges que es pot pensar com un codificador seguit d’un descodificador. Per tant, la xarxa neuronal a entrenar té dues parts, de la qual s’han provat diferents *encoders*, és a dir, classificadors. A més, a l’hora d’entrenar el model, un aspecte que té influència és la *loss function*, ja que és la que guia el procés d’entrenament de la xarxa. Així doncs, hi ha diferents opcions tant de codificadors com de funcions de pèrdua que s’han utilitzat durant aquest projecte i que han estat considerats per trobar el millor model.

Taula 2. Percentatge de píxels i pesos de cada classe en les dades d'entrenament.

	Tumor	Estroma	Infiltració inflamatòria	Necrosi	Altres
Percentatge	0.4112	0.3582	0.1116	0.0691	0.0499
Pesos	0.0499	0.0573	0.1840	0.2972	0.4116

Pel que fa als codificadors, els models que s'han entrenat han fet servir alguns dels diferents *encoders* introduïts a l'apartat 3.3.1, concretament ResNext, DenseNet, Xception i EfficientNet. En el cas d'aquest projecte, els pesos utilitzats per inicialitzar l'*encoder* s'obtenen del model corresponent entrenat amb Imagenet. És a dir, pels diferents models classificadors utilitzats, la inicialització dels pesos en cada cas s'ha fet segons l'entrenament amb Imagenet amb l'arquitectura corresponent.

Respecte la funció de pèrdua, les dades utilitzades de BCSS estan força desbalancejades. Així doncs, es considera que utilitzar la *Weighted Cross Entropy Loss* és una bona manera d'evitar que la xarxa es fixi molt en les classes més freqüents. A la Taula 2 es pot observar el percentatge de píxels de cada classe en les imatges d'entrenament i els pesos corresponents a cadascun dels *labels*, calculats a partir del percentatge utilitzant la Eq.(10). Finalment, però, els valors utilitzats realment com a pesos s'obtenen a partir de la normalització dels valors dels pesos obtinguts anteriorment.

$$w_i = \frac{1}{p_i} \quad (10)$$

4.3. Preparació dades de l'ICS

El model s'entrena utilitzant les dades del *Breast Cancer Semantic Segmentation dataset*, però l'objectiu del projecte és fer prediccions en les imatges de l'ICS. Per tant, un cop obtingut el model entrenat optimitzant els paràmetres amb les dades de BCSS, la idea és fer inferència en les *Whole Slide Images* de l'Institut Català de la Salut.

4.3.1. Creació dels *patches*

Per fer això, s'han d'obtenir *patches* de les WSI tal que continguin teixit per tal de poder passar-los com entrada de la xarxa i obtenir les prediccions. Per aquest procés, un fet a tenir en compte és la magnificació de les dades, ja que el model ha estat entrenat amb una certa magnificació. Per tant, el *patches* a extreure de les imatges de l'ICS han de ser de la mateixa magnificació, han de tenir el mateix nivell de detall per tal d'obtenir resultats òptims.

Per crear els *patches* de zones que contenen la mostra, el primer que s'ha fet és una classificació dels píxels per tal de diferenciar les regions amb i sense teixit. Per fer això, s'ha utilitzat la WSI amb poca resolució i, per tant, de mida més petita per tal de poder treballar amb tota la imatge sencera sense problemes de memòria.

$$\|(p - w)\| \geq th \quad (11)$$

El mètode utilitzat per fer aquesta binarització es basa en la diferenciació entre els píxels de teixit i de fons mitjançant la comparació entre el valor del píxel amb el color blanc pur (ja que és el color del fons teòricament). Per tant, es mira si aquesta diferència és més gran que un *threshold* per tal de decidir si el píxel correspon a un tros de teixit o del fons, aplicant la Eq.(11). A partir de la màscara binària obtinguda, es pot fer una aproximació de la *bounding box* que conté la mostra i, d'aquesta manera, obtenir les coordenades d'inici i de final de la "zona de teixit", tal i com es representa a la Figura 19.

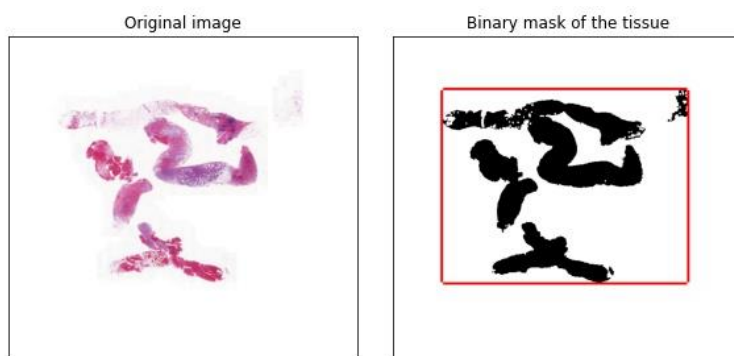


Figura 19. Màscara binària del teixit.

A continuació, es fa un "escaneig" per tota la *bounding box* de manera que es passi per totes les zones del seu interior. Aquest pas es fa a la magnificació òptima, és a dir, a la mateixa amb la qual s'ha obtingut el millor model entrenat amb les dades BCSS. Per tant, un cop obtinguda la capsula englobant, es troba una aproximació a la magnificació desitjada utilitzant els factors de *downsample* per tal d'obtenir els valors corresponents de les cantonades. Així doncs, s'obtenen *patches* de mida 5.000 × 5.000 pertanyents a regions consecutives. En cadascun d'aquests, per tal d'evitar trossos sense teixit, s'ha aplicat un altre cop la Eq.(11) per eliminar els *patches* que continguin poc teixit o que només continguin fons. És a dir, es torna a fer la mateixa diferenciació i es mira si la quantitat de píxels de teixit és suficient o no. Per tant, els *patches* dels quals s'obtindrà la predicció són aquells amb suficient teixit, per tal d'evitar buscar zones tumorals en regions on no hi ha gaire tros de mostra. A la Figura 20 es poden observar alguns exemples dels *patches* resultants per diferents imatges.

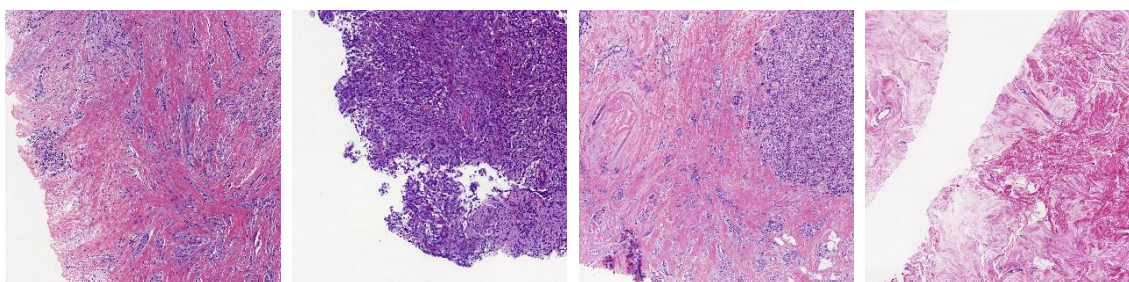


Figura 20. Patches de les imatges de l'ICS amb suficient teixit.

Així doncs, un cop obtinguts els diferents *patches* d'una *Whole Slide Image*, l'objectiu és utilitzar el model entrenat per tal de fer inferència en cadascuna de les imatges extretes. D'aquesta manera s'obtenen les màscares amb els diferents elements predits per la xarxa en cadascun dels *patches*. Així doncs, per una WSI concreta, el resultat obtingut són les màscares de cadascuna de les regions amb teixit.

4.3.2. Eina de visualització

Per tal de visualitzar les dades conjuntament amb les anotacions, s'ha fet servir una eina anomenada *Digital Slide Archive* (DSA) [47], que és una plataforma per emmagatzemar, analitzar, gestionar, visualitzar i anotar grans *datasets* d'imatges. El DSA consta d'una eina d'anàlisi (*HistomicsTK*), una interfície per visualitzar les mostres i gestionar les anotacions (*HistomicsUI*), una base de dades (*Mongo*), i un servidor web que proporciona una API i eines per gestionar les imatges (*Girder*).

En el nostre cas, el que s'ha utilitzat és *HistomicsUI* per tal de poder visualitzar i treballar amb les anotacions. Aquesta és una aplicació que permet examinar, anotar i processar imatges histològiques per extreure elements a baixa i alta resolució. A la Figura 21 es pot veure com és l'eina i la visualització d'algunes anotacions d'una WSI del *dataset* BCSS.

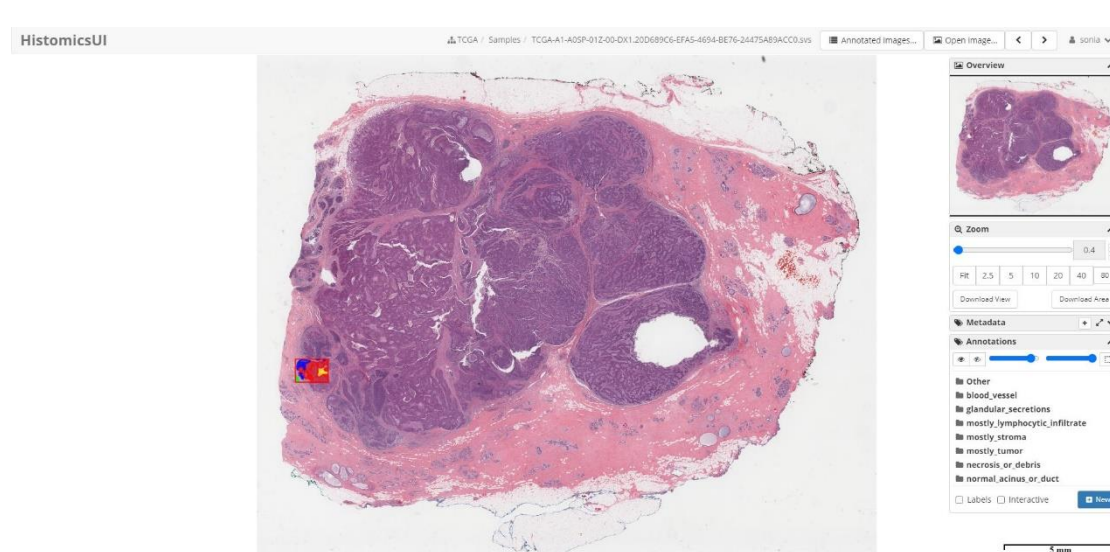


Figura 21. Digital Slide Archive.

4.4. Registre

Paral·lelament a aquest projecte, s'ha realitzat un altre treball conjuntament amb el grup de DigiPatCIS corresponent a l'assignatura *Introduction to Research* i que, per tant, estan relacionats. Aquest consisteix en fer el registre de diverses imatges histològiques del mateix pacient però de diferents tincions. L'objectiu és alinear una àrea concreta del teixit en dues tincions diferents de manera que es pugui fer un anàlisi dels elements de les dues imatges i combinar la informació obtinguda en cadascuna de les tincions per tal de fer un diagnòstic més precís. A la Figura 22 es pot veure un exemple del registre a baixa resolució, on la primera imatge correspon a la tinció de referència, la segona al teixit d'una altra tinció i que es vol alinear i, la última és el resultat després del registre.

Com he esmentat, aquests dos projectes estan relacionats, ja que es vol aplicar el registre a les zones tumorals obtingudes a partir de la predicció en la tinció Hematoxilina-Eosina per obtenir la mateixa regió en una altra tinció. D'aquesta manera, es pot combinar la informació extreta en cadascuna d'elles i fer un diagnòstic més precís i complet.

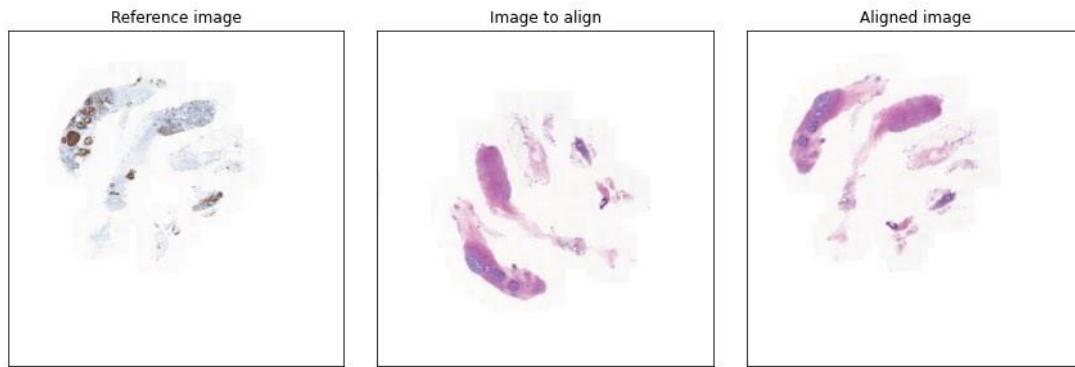


Figura 22. Registre de dues tincions diferents a baixa resolució.

Aquest registre, però, no es fa només a nivells de baixa resolució, sinó que també es pot aplicar en zones amb molt nivell de detall, per exemple, a nivell de cèl·lula. En aquest cas, la idea és seleccionar una regió en una de les tincions i registrar la mateixa zona de l'altra *Whole Slide Image*, tal i com es pot veure a la Figura 23. D'aquesta manera, es pot fer un anàlisi a partir de les dues imatges i combinar la informació aportada per cadascuna de les tincions en la zona per fer un diagnòstic més precís de la possible zona tumoral. Aquest anàlisi pot ser tant visual per part dels patòlegs com automàtic a partir de xarxes neuronals, a les quals se'ls pot passar com entrada les dues imatges alineades per tal de que aprenguin millor a partir de la informació de les dues tincions.

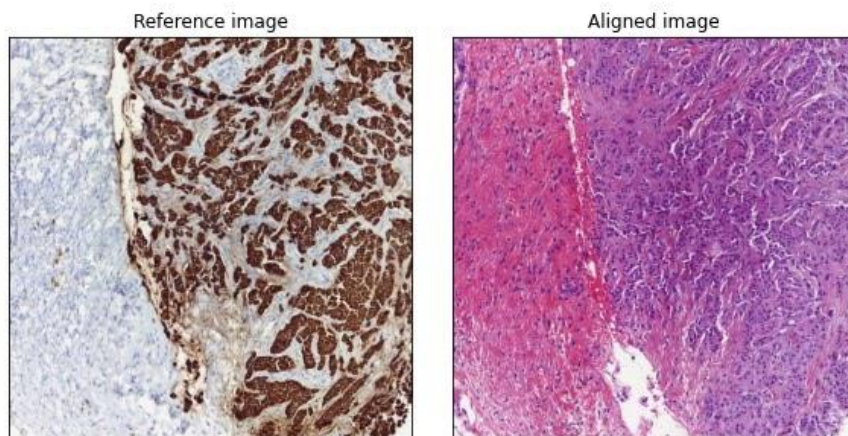


Figura 23. Registre de dues imatges a alta resolució.

A continuació, s'explica l'algorisme del registre, el qual consta de tres etapes diferents precedit per un primer pas imprescindible per donar robustesa a l'algorisme. A la Figura 24 es pot observar un esquema de cadascun dels passos, on els blocs que es troben dins del rectangle verd fan referència als 3 passos principals de l'algorisme de registre.

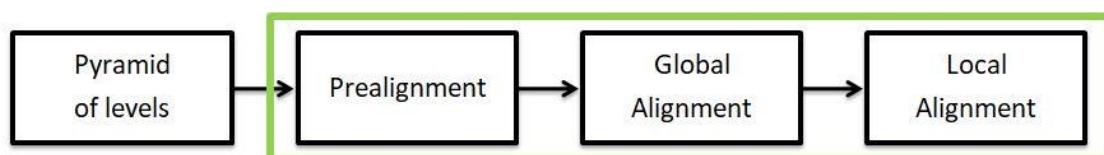


Figura 24. Esquema de l'algorisme de registre.

Per tant, el primer pas és construir una piràmide dels diferents nivells. La motivació d'aquest és que, a partir de la imatge de referència de la Figura 23, no és possible trobar la regió en l'altra tinció directament. Per tant, la idea és ampliar progressivament el camp visual de manera que s'acabi obtenint la imatge amb el teixit complet a baixa resolució. Així doncs, l'algorisme utilitzarà en les diferents etapes els camps visuals intermedis obtinguts amb la piràmide de nivells per tal d'actualitzar els paràmetres calculats en els anteriors nivells i així millorar l'alineament iterativament.

Un cop fet això, la idea és fer una primera aproximació a partir dels nivells de baixa resolució en els quals és visible tot el teixit, com a la Figura 22. En aquest pas, ens centrem en la forma de la mostra, així que es treballa a partir d'una màscara binària de la zona de teixit. Aquesta màscara s'obté utilitzant el mateix procediment explicat a l'apartat anterior, és a dir, comparant el valor del píxel amb el color blanc. A partir d'aquesta, es calculen els paràmetres òptims del vector de translació i de l'angle de rotació per tal d'alinear les imatges.

A continuació, en nivells de resolució intermedis, es fa un registre global. En aquests cassos, no és visible tota la mostra, però tot i això, ens centrem encara en la forma del teixit per tal de buscar la transformació que maximitzi la similitud entre les dues imatges. Així doncs, es torna a treballar a partir de la binarització del teixit i, a més, es calcula una funció distància per tal d'utilitzar una funció més suau. A partir d'aquesta, com en el cas anterior, s'obtenen el vector de translació i l'angle de rotació òptims.

Finalment, a nivells d'alta resolució, es fa un alineament local, en el qual ens centrem en l'estructura interna del teixit, és a dir, en els elements presents en la regió. Degut a què les imatges original són molt diferents, el que es fa en aquest cas és buscar una manera de fer-les més similars. Així doncs, el que s'ha fet és extreure la component Hematoxilina de cadascuna de les tincions per tal d'obtenir els nuclis ressaltats en cadascuna d'elles. A partir d'aquestes imatges, es calculen els paràmetres necessaris per tal de definir la transformació òptima per fer el registre amb molt nivell de detall.

Així doncs, els resultat obtingut són dues imatges corresponents a la mateixa regió, però de dues tincions diferents, com a la Figura 23. Per tant, la idea és que, a partir de la predicció de zones tumorals en la tinció HE, es pugui aplicar aquest registre per obtenir la mateixa regió amb una altra informació destacada, és a dir, en una altra tinció. D'aquesta manera, es pot combinar la informació extreta en cadascuna d'elles i fer un diagnòstic més precís i complet. En el cas del nostre projecte, això significaria buscar en les imatges d'altres tincions les àrees corresponents a les zones tumorals detectades per complementar l'anàlisi sobre Hematoxilina-Eosina amb informació de proliferació, presència d'epiteli o de reacció a estrògens o progesterona d'aquella zona, que permet fer un diagnòstic més precís i complet.

5. Resultats

En aquest apartat s'especifiquen les diferents proves referents a la solució proposada utilitzant el BCSS *dataset*, els resultats obtinguts al fer inferència amb les imatges de l'ICS i la seva aplicació amb el registre. Per tant, s'expliquen els resultats obtinguts segons els diferents elements explicats en l'apartat anterior (*encoders*, *loss function*, etc.) fent èmfasi en el millor model i es mostren els resultats obtinguts tant quantitativa com qualitativament. A continuació, es mostra el resultat de les prediccions de les imatges de l'Institut Català de la Salut a partir del model obtingut amb les de BCSS. Finalment, es mostra un exemple de l'aplicació del registre relacionat amb aquest projecte utilitzant les prediccions obtingudes.

5.1. Resultats del BCSS dataset

Primer de tot, s'entrenen diferents models a partir de les dades del *Breast Cancer Semantic Segmentation* (BCSS) per tal d'obtenir l'òptim, és a dir, aquell que aprèn millor a fer la segmentació de les imatges d'Hematoxilina-Eosina per tal de detectar càncer de mama. Com s'ha explicat a l'apartat 4.1, les imatges s'han dividit en tres conjunts: entrenament, validació i test. Així doncs, s'entrena la xarxa a partir del primer i s'utilitzen les imatges del segon grup per tal de triar el millor model en cada cas. Finalment, un cop obtingut el model òptim, s'aplica a les imatges del conjunt de test per tal d'obtenir els resultats amb imatges que són completament noves i la xarxa no ha vist mai.

A l'hora de triar el millor model, el que s'ha fet és calcular les mètriques per separat per cadascuna de les classes i , a continuació, calcular el valor mitjà de cada mètriques per obtenir el valor del model en cada cas. Per exemple, pel que fa al *recall*, el que es fa és calcular a partir de la matriu de confusió el seu valor per cada classe i , a continuació, fer la mitjana dels cinc *recalls* per tal d'obtenir el valor mitjà, el del model en general. D'aquesta manera, es busca un equilibri entre totes les classes, i s'intenten eliminar problemes derivats de les classes desbalancejades. Per tant, es busca que el model sigui bo per totes les classes. En el nostre cas, la detecció de tumors, creiem que és molt important que la xarxa no tingui falsos negatius, ja que el cost associat a aquests és molt gran degut a les conseqüències derivades. És a dir, no volem que la xarxa no predigui zones que sí són tumorals pels efectes que pot tenir. Per tant, ens centrem en el *recall* a l'hora d'escollir el model òptim, és a dir, escollim el millor conjunt de pesos de la xarxa a partir del màxim *recall* obtingut en el conjunt de validació.

Per triar el millor model s'han d'optimitzar diversos hiperparàmetres que afecten el rendiment de la xarxa, a la manera d'aprendre de les dades, com per exemple, la mida del *batch*, el valor del *learning rate*, etc. A més, hi ha altres factors que són molt importants per tal d'obtenir bons resultats com els *encoders* i la *loss function*.

En aquest projecte també s'ha de tenir en compte un aspecte referent a les imatges, la magnificació. Com s'ha esmentat anteriorment, les imatges del BCSS *dataset* poden descarregar-se amb diferents nivell de detall. Així doncs, s'han provat diferents magnificacions per tal de veure amb quina d'aquestes la xarxa té més capacitat d'aprendre a segmentar les classes. En aquest cas, s'han provat 3 magnificacions diferents: 10, 20 i 40, sent aquesta última l'original, és a dir, amb la que s'ha capturat el teixit i , per tant, la de màxima resolució.

A partir de les mètriques obtingudes en cada cas, els millors resultats s'han obtingut amb la magnificació 20. Per tant, es pot concloure que no és aconsellable aprendre a partir de la imatge amb tant nivell de detall com en el cas de magnificació 40, però tampoc a poca resolució com en el cas de 10. Això ens indica que la xarxa necessita tenir una visió més completa del context per fer la segmentació més correctament, és a dir, no necessita tant nivell de detall, sinó que necessita conèixer una mica més els elements propers per tal de prendre una decisió més acurada. Però també es pot concloure que la xarxa ha de poder observar certs detalls de les components fines de la imatge per fer una millor detecció.

Així doncs, les dades a utilitzar són les de magnificació 20, ja que són amb les que s'obtenen millors resultats. En aquest cas, les imatges dels conjunts d'entrenament i de validació s'obtenen com s'ha especificat a l'apartat 4.1. Per tant, aquestes són de mida 1.000×1.000 , mentre que les imatges de test no es modifiquen. Això indica que la xarxa s'entrenarà amb un conjunt d'entrenament format per 462 imatges, es validarà utilitzant 101 *patches* i, finalment, s'aplicarà a les 22 imatges de test originals.

Respecte la *loss function*, tal i com s'ha esmentat a l'apartat 3.3.1, en aquest treball s'han provat diferents funcions per a la configuració del model. A partir dels resultats hem vist que clarament la *Weighted Cross Entropy Loss* és la que ha donat millors resultats. Per tant, aquesta funció ha sigut la utilitzada en la majoria dels models, inclòs l'òptim. Un fet a destacar és que, tant la Jaccard com de la Dice *loss* han estat descartades aviat, ja que els valors de la funció de pèrdues no disminuïen molt, és a dir, el model no aprenia gaire. Aquest fet es pot observar a la Figura 25, on es mostra l'evolució de cadascuna de les funcions de pèrdua. Conseqüentment, els resultats obtinguts eren dolents i no milloraven massa en les diferents *epochs*. En canvi, en el cas de la *Weighted Cross Entropy Loss*, sí que s'ha pogut observar un comportament típic, en el qual el valor de la funció disminueix força al principi i, després, ho fa més lentament.

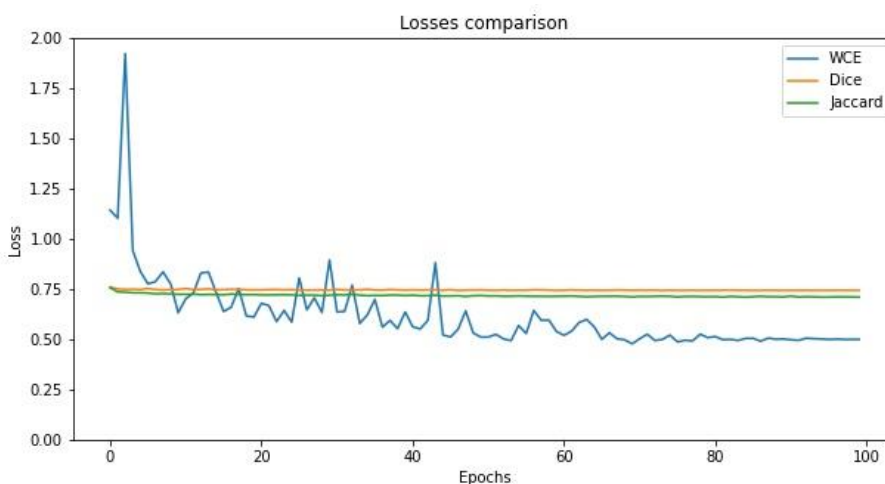


Figura 25. Comparació de cadascuna de les *loss functions*.

Un altre aspecte a tenir en compte és l'*encoder* utilitzat en la U-Net per tal d'extreure les *features* importants de les imatges per fer la segmentació. Els models que s'han entrenat han fet servir alguns dels *encoders* introduïts anteriorment, concretament ResNext, DenseNet, Xception i EfficientNet. En aquest cas, els resultats obtinguts amb cadascun dels codificadors han estat força semblants, no s'ha vist cap variació remarcable respecte els diferents models classificadors. Tot i així, s'ha pogut apreciar

que eren una mica millors els valors en el cas de la ResNext i de la EfficientNet, tal i com es mostra a la Taula 3. Per aquest motiu, aquestes dues arquitectures d'*encoder* han sigut les que s'han utilitzat per fer la resta d'optimitzacions.

Taula 3. Resultats obtinguts amb els diferents encoders.

	ResNext	DenseNet	Xception	EfficientNet
Recall	0.7762	0.7367	0.7601	0.7966
Precision	0.7072	0.6576	0.6646	0.7407
F-Score	0.7323	0.6824	0.6850	0.7622
Accuracy	0.8795	0.8424	0.8455	0.8969

Respecte els altres hiperparàmetres, s'han provat diferents valors de *batch size*, *learning rate*, número d'*epochs* i de la mida de les imatges d'entrada (és a dir, del resultat al aplicar el *crop*). A partir de les diferents combinacions, s'han obtingut els millor resultats entrenant el model amb imatges de mida 512×512 durant 100 *epochs*, fent servir un *batch* de mida 8 i un *learning rate scheduler* del tipus *Cosine Annealing* [50], [51] amb *learning rate* 0.00068. Aquest consisteix a fer decreixer el *learning rate* simulant la meitat d'una corba de cosinus, tal i com es pot veure a la Figura 26. Per tant, es comença amb un *learning rate* que decreix la *loss function* a un mínim i, gradualment, es va disminuint el valor del *learning rate* quan ens apropem al mínim de la funció de pèrdua.

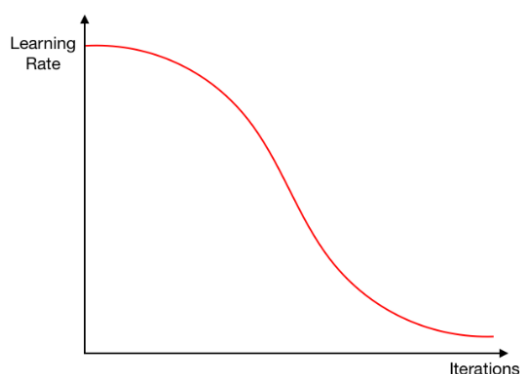


Figura 26. Evolució del learning rate en el Cosine Annealing scheduler.

Així doncs, el model òptim i que, per tant, s'ha utilitzat per fer inferència de les dades de l'ICS té una arquitectura de la U-Net fent servir com *encoder* una EfficientNet. Per tal d'optimitzar els pesos de la xarxa, s'ha utilitzat la *Weighted Cross Entropy Loss* amb els pesos especificats a la Taula 2 per cadascuna de les classes. Finalment, cal destacar que la magnificació òptima, és a dir, amb la que la xarxa fa la millor segmentació és de 20. Per tant, el model ha estat entrenat amb els següents paràmetres:

- Imatges de magnificació 20
- EfficientNet com model classificador, és a dir, *encoder*
- *Weighted Cross Entropy Loss* com a funció de pèrdua
- 100 *epochs*
- *Batches* de mida 8
- Imatges de 512×512 píxels
- *Learning rate scheduler* del tipus *Cosine Annealing*.

Durant l'entrenament de la xarxa neuronal, s'obtenen diferents mètriques en cadascuna de les *epochs*, tant amb les dades del conjunt de *train* com en les de validació. D'aquesta manera, es pot tenir una idea de com ha sigut l'entrenament i veure, per exemple, aspectes com si s'ha produït sobreentrenament o no. Concretament, en aquest document es mostren quatre gràfiques referents a les mètriques obtingudes durant el procés d'aprenentatge de la xarxa: el valor de la *loss function*, del *recall*, de la precisió i de l'F-Score.

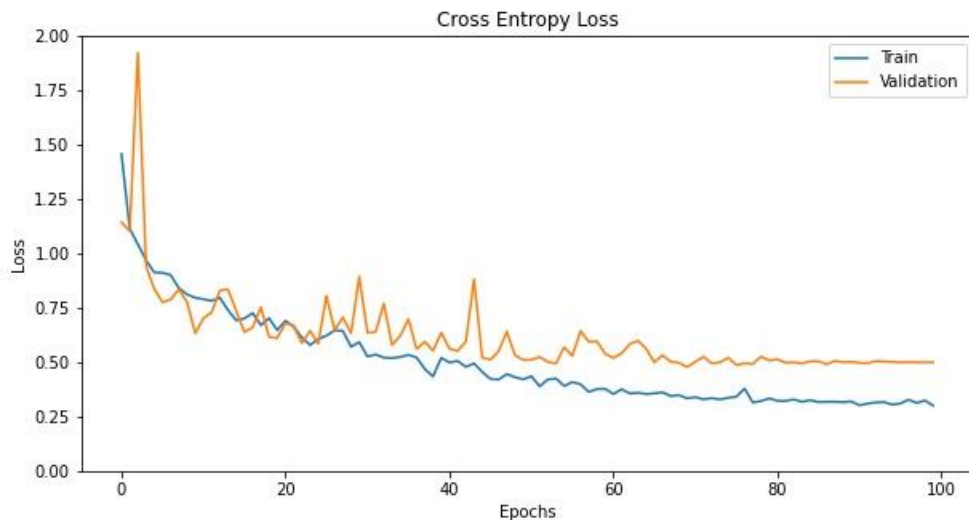


Figura 27. Weighted Cross Entropy Loss.

A la Figura 27 es pot veure l'evolució del valor de la funció de pèrdua tant en el conjunt d'entrenament com en el de validació. Tal i com s'ha esmentat anteriorment, es pot veure un comportament més o menys típic d'aquestes funcions, ja que al principi el valor cau de manera força ràpida i, a continuació, comença a disminuir de forma més lenta. Per tant, es pot dir que la xarxa aprèn bé degut a que aquest valor és cada cop més petit, i sembla que es produeix una mica d'*overfitting*, però no gaire.

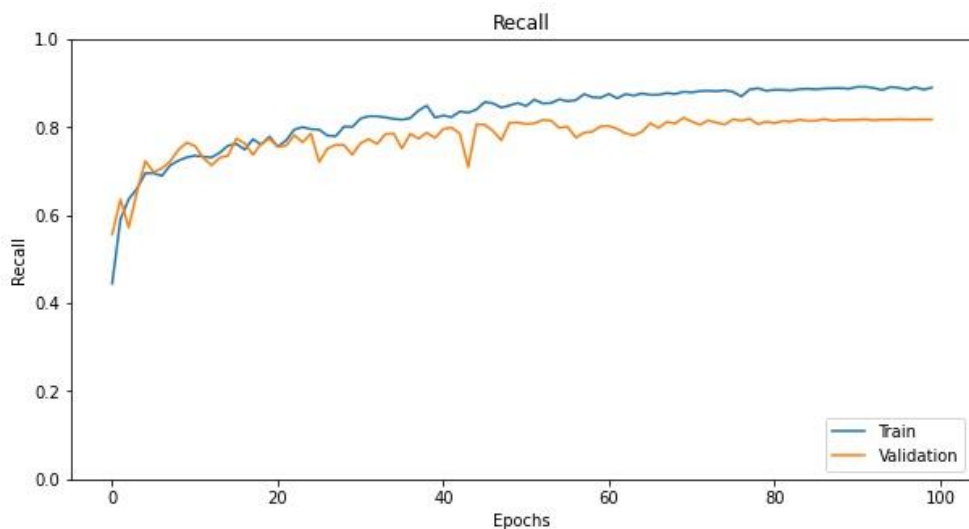


Figura 28. Recall.

Pel que fa el valor del *recall*, podem veure a la Figura 28 el mateix que en el cas anterior, és a dir, que la xarxa aprèn a segmentar les classes. Això és degut a que aquest valor augmenta amb el pas de les *epochs*. Com en el cas de la *loss function*,

aquest creixement no és constant, ja que al principi el valor del *recall* augmenta de forma més ràpida i després ho fa lentament. En aquesta gràfica també es pot apreciar que la xarxa fa una mica d'*overfitting*, degut a la diferència entre els valors obtinguts amb les dades d'entrenament i de validació.

Pel que fa la Figura 29, es pot veure l'evolució de la precisió durant el procés d'entrenament del model. Aquest cas té la mateixa tendència que el *recall*, és a dir, el valor comença creixent força i després ho va fent d'una manera més lenta. Un aspecte a destacar és que, si comparem la Figura 28 i la Figura 29, es pot veure que el valor de la precisió és sempre menor que el del *recall*. Aquest fet ens afavoreix, ja que estem interessats en valors de *recall* alts per tal de no tenir gaires falsos negatius, tal i com s'ha esmentat anteriorment.

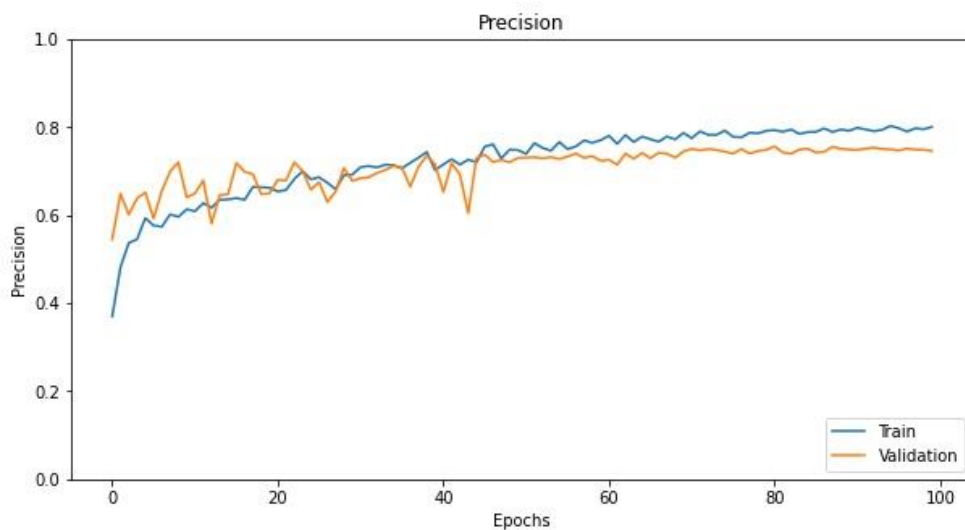


Figura 29. Precisió.

Finalment, a la Figura 30, es mostra el valor de F-Score obtingut en cadascun de les *epochs*. Tal i com s'ha explicat a l'apartat 3.3.1, aquesta mètrica s'utilitza per balancejar els valors obtinguts en cada cas. Així doncs, la tendència de la gràfica és la mateixa que en els dos casos i els valors mostrats representen una mitjana harmònica entre el *recall* i la precisió del model.

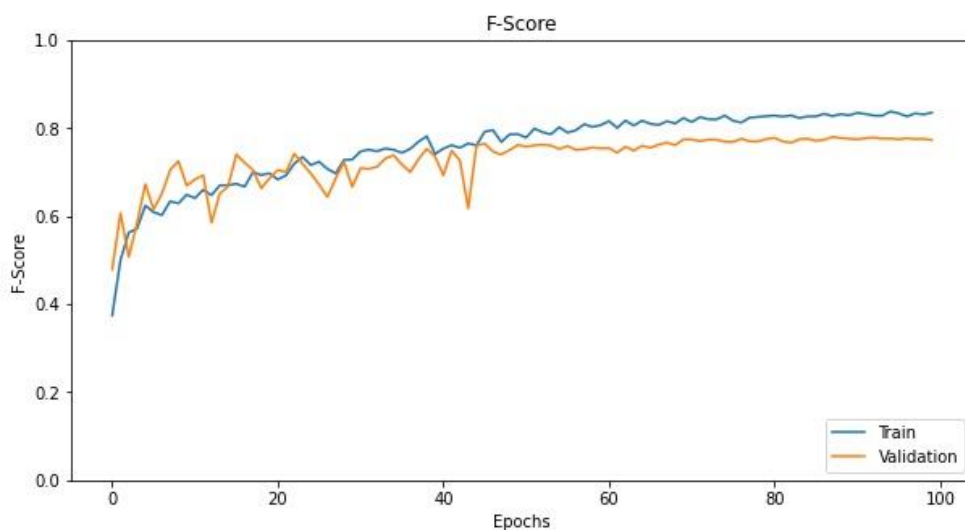


Figura 30. F-Score.

Un cop s'ha entrenat la xarxa i s'ha obtingut el model òptim, aquell que ens ha donat un valor de *recall* més alt en les dades de validació, s'han calculat diferents mètriques derivades, tant en el cas de les imatges d'entrenament com en les de validació i test. Així doncs, s'han calculat els valors de les matrius de confusió i, a partir d'aquestes, les mètriques de *recall*, precisió, F-Score i *accuracy* de cadascuna de les classes individualment.

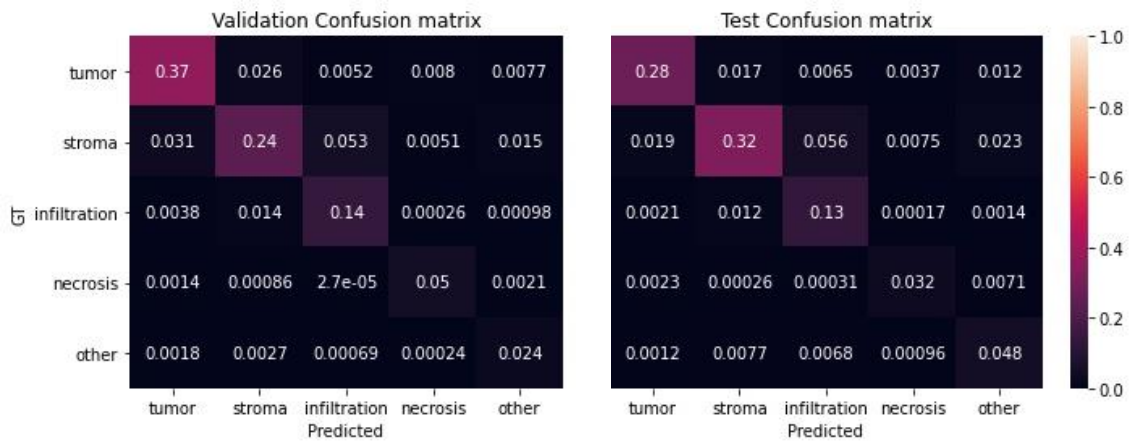


Figura 31. Matrius de confusió de les dades de validació i test.

Les matrius de confusió s'utilitzen per avaluar el model comparant els valors anotats (els del *ground truth*) amb les prediccions. En el nostre cas, les columnes d'aquestes matrius representen els valors predits pel nostre model entrenat, mentre que les files corresponen a les anotacions, a la classe correcta. Per tant, en cadascuna de les caselles de la matriu, el valor a_{ij} indica la quantitat d'observacions anotades amb el *label i* que han estat classificades com a classe *j*. A la Figura 31 es pot observar aquesta matriu en el cas de les dades de validació i test, amb el peruntatge corresponent a cadascuna de les caselles.

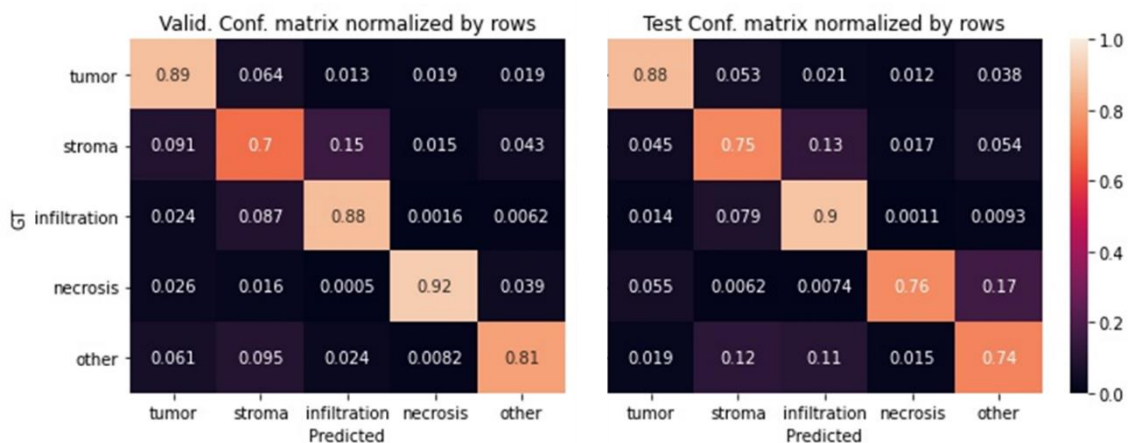


Figura 32. Matrius de confusió normalitzades per files de les dades de validació i test.

Per obtenir informació més detallada per cadascuna de les classes, el que es fa és normalitzar aquestes matrius per files i/o per columnes. En el primer cas, es pot veure, per cadascuna de les observacions anotades, la proporció amb la que el model prediu cadascun dels *labels*. En canvi, si es mira la matriu de confusió normalitzada per columnes, el que es pot analitzar és, per cadascuna de les prediccions, la proporció respecte el *ground truth*. Per tant, al normalitzar per files, la diagonal indica el valor del

recall de cada classe, mentre que fer-lo per columnes fa que s'obtinguin valors corresponents a la precisió. A la Figura 32 es poden veure les matrius de confusió normalitzades per files tant de les dades de validació com les de test, ja que els valors representats estan relacionats amb el *recall*.

A partir dels valors mostrats a la Figura 32 es pot veure que no hi ha gran diferència entre els resultats obtinguts amb les dades de validació i les de test. Les classes on hi ha més diferència són en el cas de la necrosi i d'altres elements, tot i que els resultats no són dolents en les imatges de test. Pel que fa la classe tumor i infiltració inflamatòria, es pot veure que el *recall* obtingut és molt alt, gairebé del 0.9. Per últim, l'estroma té una valor de 0.7 en els dos casos, així que també es pot considerar que la predicció és bona.

A partir d'aquestes matrius de confusió, es poden calcular les mètriques de *recall*, precisió, F-Score i *accuracy* per cadascuna de les classes i, a partir d'aquests, el valor mitjà de cadascuna d'elles. La Taula 4 mostra els valors obtinguts en el cas del *dataset* de validació per cadascuna d'aquestes mètriques. En canvi, pel que fa el conjunt d'imatges de test, es mostren de manera detallada els resultats de F-Score i d'*accuracy* a la Taula 5.

Taula 4. Mètriques de les dades de validació per cada classe i la mitjana.

	Tumor	Estroma	Infiltració inflamatòria	Necrosi	Altres	Mitjana
Recall	0.8856	0.6974	0.8812	0.9186	0.8123	0.8390
Precision	0.9051	0.8468	0.7006	0.7863	0.4763	0.7430
F-Score	0.8952	0.7649	0.7806	0.8473	0.6005	0.7777
Accuracy	0.8953	0.8189	0.9052	0.9780	0.9618	0.9119

Taula 5. Mètriques de les dades de test per cada classe i la mitjana.

	Tumor	Estroma	Infiltració inflamatòria	Necrosi	Altres	Mitjana
F-Score	0.8962	0.8191	0.7607	0.7387	0.6135	0.7657
Accuracy	0.9214	0.8249	0.8956	0.9725	0.9254	0.9080

A partir dels valors de la Taula 4 es pot veure que el *recall* en la majoria dels cassos és força alt. És important tenir en compte que en la classe relativa als tumors, tant el valor de *recall* com el de la precisió és molt alt, tot i que aquesta última mètrica és la que té un valor més gran. També s'ha de dir que la classe en la que hi ha més diferència entre la precisió i el *recall* és "Altres". Això pot ser degut a que és una classe en la que s'han agrupat moltes altres diferents (com es pot veure a la Taula 1), fet que fa que sigui més difícil de predir correctament. Respecte els valors de la Taula 5, es pot veure que la mètrica F-Score té uns valors força semblants als de les dades de validació per cadascuna de les classes. En el cas de l'estroma, millora degut a que el *recall* aconseguit amb les imatges de test és superior, de 0.7524. En canvi, en el cas de la necrosi, el valor és menor en aquest cas perquè el *recall* disminueix molt, tal i com es pot veure també en les matrius de confusió de la Figura 32.

Finalment, es mostren alguns exemples qualitius dels resultats obtinguts en el cas de les imatges del conjunt de test. Per tal de mostrar aquestes imatges i de poder

comparar-les amb els valors reals, primer es mostra la imatge original, a continuació la màscara del *ground truth* i, finalment, la màscara predita pel model seleccionat. Per tant, a partir dels resultats presentats a la Figura 33, es pot veure que la xarxa prediu força bé les zones on estan presents cadascuna de les classes. Tot i això, sí que sembla que hi ha casos en els que la segmentació no acaba de ser del tot precisa, ja que la delimitació de la zona afectada no és del tot igual. A més, en les imatges es pot veure qualitativament el que ja s'havia vist a partir de les mètriques calculades anteriorment, com el fet de què es prediu a vegades necrosi que no n'hi ha, o la classe "Altres" que té un comportament semblant.

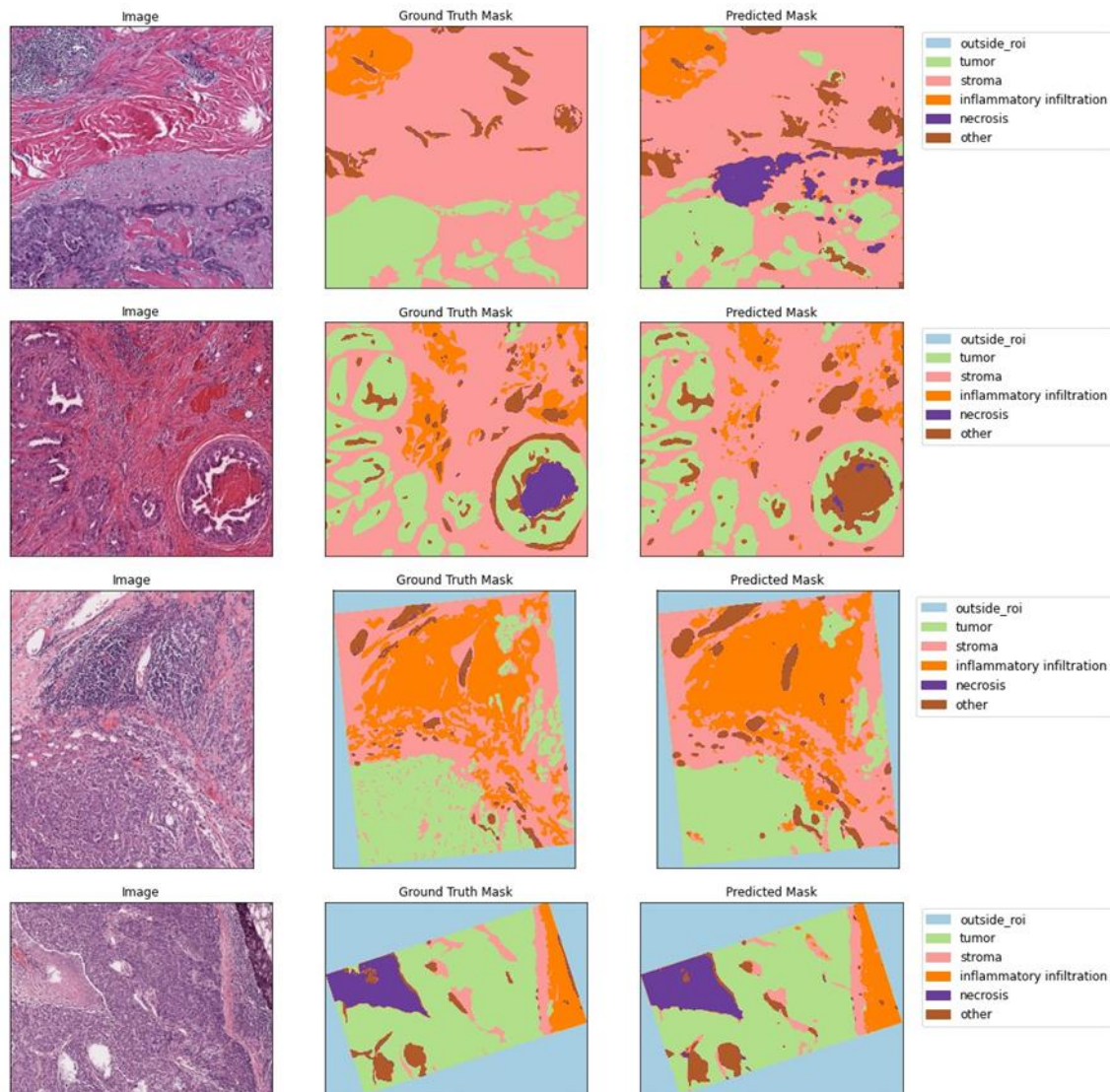


Figura 33. Prediccions imatges de test del BCSS dataset.

5.2. Inferència dades de l'ICS

L'objectiu d'aquest projecte és entrenar un model per tal de fer prediccions en les imatges de l'ICS. Per tant, un cop obtingut el millor model a partir de les dades del *Breast Cancer Semantic Segmentation (BCSS)*, es fa inferència amb les imatges de l'Institut Català de la Salut. Tal i com s'ha explicat anteriorment a l'apartat 4.3, el que

es fa és obtenir *patches* de les *Whole Slide Images* de l'ICS i fer la predicció d'aquests utilitzant el millor model entrenat anteriorment.

En aquest cas, però, no hi ha *ground truth*, així que no es poden obtenir mètriques quantitatives. Per tal de fer una valoració dels resultats, només es poden fer valoracions qualitatives des del nostre punt de vista, ja que no ha sigut possible ensenyar aquestes imatges als patòlegs per tal d'obtenir les seves impressions. Tot i això, a partir dels resultats obtinguts en les imatges de la Figura 34, sembla que el model detecta bé les zones tumorals en les imatges de l'ICS. A més, es pot veure que el model és robust a la variabilitat de les dades, ja que ha estat entrenat imatges molt diferents (tal i com es pot veure a la Figura 15 i a la Figura 16), i tot i així els resultats obtinguts són bons.

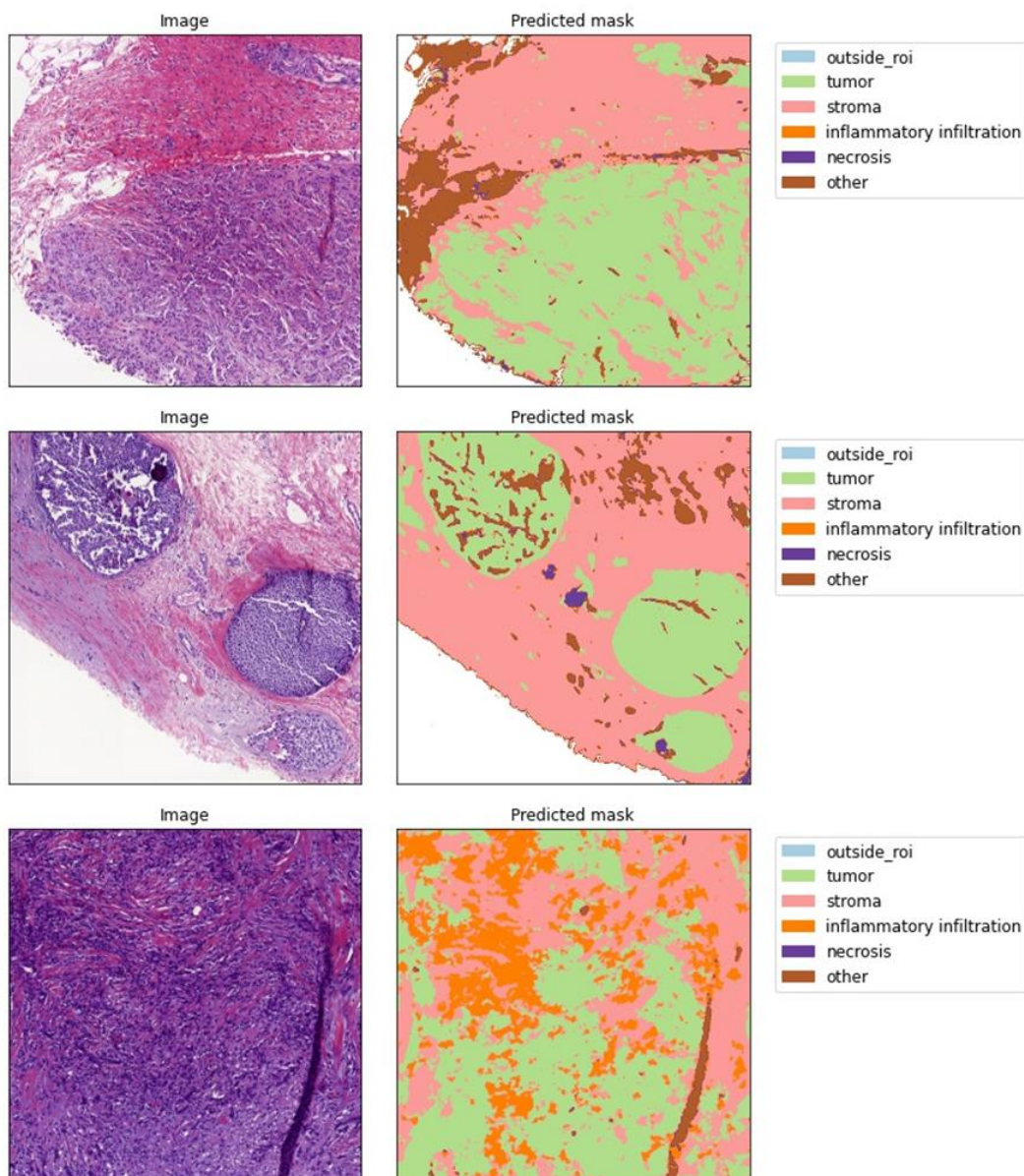


Figura 34. Prediccions imatges de l'ICS.

La gran utilitat d'aquestes prediccions, però, no és veure la màscara obtinguda de la segmentació per cadascun dels *patches* extrets de les WSI. El que és més interessant pels patòlegs és poder visualitzar les prediccions en un entorn en el qual es pugui

veure tant el teixit com la segmentació i que, a més, permeti navegar per la imatge per tal d'analitzar diferents parts de la mostra. Això s'ha pogut aconseguir gràcies a l'aplicació Digital Slide Archive (DSA), la qual s'ha explicat anteriorment a l'apartat 4.3.2 i que permet visualitzar la imatge i les anotacions a diferents resolucions de manera que els usuaris puguin navegar per tot el teixit. Així doncs, per poder visualitzar els resultats obtinguts, s'ha hagut de passar la màscara amb les prediccions de les classes (és a dir, una imatge) a un fitxer de tipus *.json* amb cadascuna de les diferents regions anotades, i pujar aquests fitxer al servidor de l'aplicació.

Tal i com es pot veure a la Figura 35, les anotacions corresponents a les prediccions del model entrenat són visibles conjuntament amb la imatge gràcies a l'aplicació. A més, es pot veure com, pujant les segmentacions de les diferents regions del teixit, es pot obtenir una detecció de les zones tumorals en gairebé tot el teixit. S'ha de tenir en compte que aquesta eina és interactiva i que, per tant, es poden visualitzar cadascuna de les classes per separat si es vol o treure la segmentació de manera que només es pugui veure la imatge original.

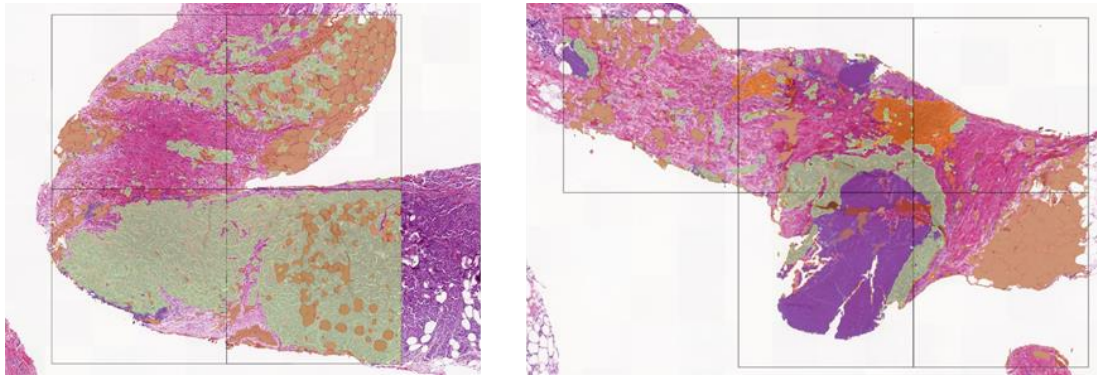


Figura 35. Exemples de les anotacions visualitzades mitjançant DSA.

5.3. Registre

Per acabar l'explicació dels resultats, es mostren alguns exemples obtinguts a partir del registre. D'aquesta manera es pot veure la necessitat d'aquesta part en futurs passos del projecte DigiPatICS, ja que serà possible entrenar models utilitzant la informació de dues o més tincions diferents.

Així doncs, a la Figura 36 es mostra un exemple de registre de la tinció HE amb alguna de les altres, CK19, KI67 i RE. Així doncs, un cop obtinguda la segmentació en la imatge d'Hematoxilina-Eosina de la Figura 34, la idea és mirar la resta de tincions en la mateixa àrea i tenir en compte la resta d'elements utilitzats per fer el diagnòstic precís. En aquest cas, per exemple, de la tinció KI67 i RE s'haurien de mirar aspectes dels nuclis cel·lulars com la intensitat de la tinció en aquesta zona i la quantitat de cèl·lules que han reaccionat.

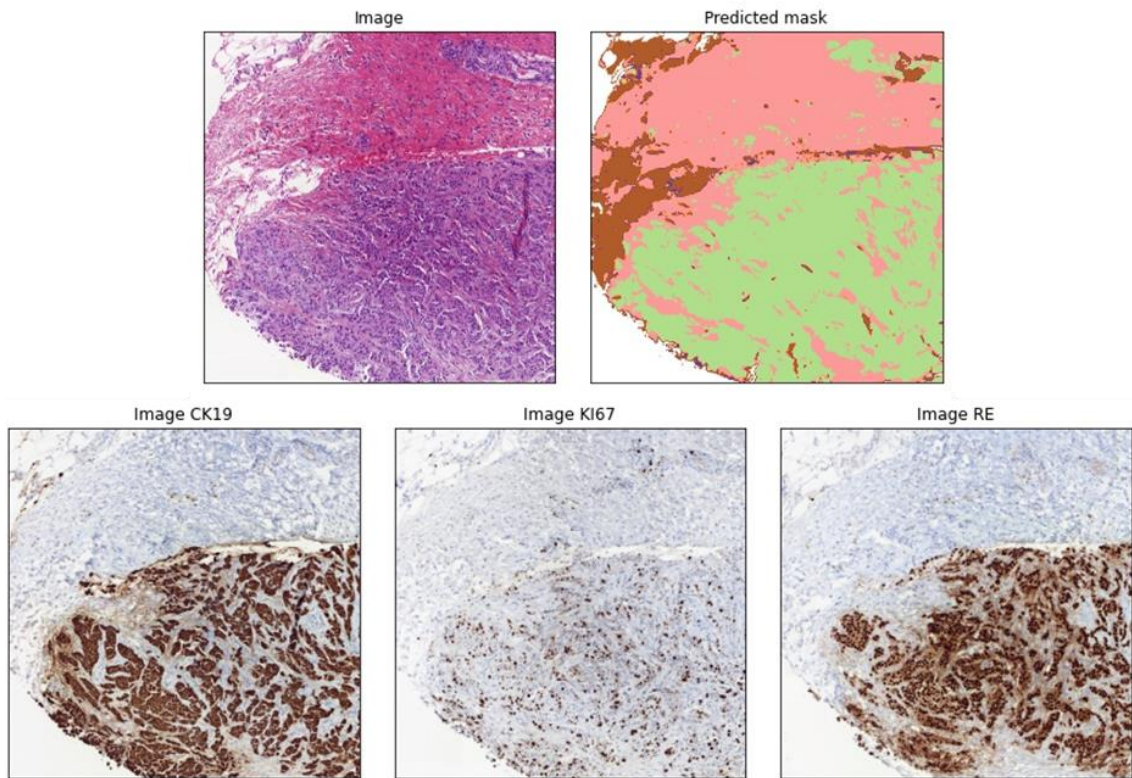


Figura 36. Registre d'una àrea de la regió HE.

Per tant, es podrien passar algunes d'aquestes imatges com entrada d'una xarxa neuronal i predir, no només les zones tumorals, sinó els diferents graus a partir de la informació obtinguda de cadascuna de les tincions i la combinació d'aquests aspectes per part del model. Això, però, és una aplicació futura d'aquest algorisme en el qual, primer, s'han de detectar les zones on hi ha cèl·lules tumorals i, a continuació, utilitzar la resta de tincions per tal d'obtenir informació addicional sobre el tumor.

6. Conclusions i treball futur

Aquest treball s'ha desenvolupat dins del projecte DigiPatICS per tal de detectar zones tumorals en imatges digitals histològiques de possibles cassos de càncer de mama utilitzant una xarxa neuronal. Concretament, es treballa amb unes imatges anomenades *Whole Slide Images* que contenen trossos de teixit tenyits per tal de destacar certs aspectes de la mostra extreta. En el cas d'aquest projecte, s'ha treballat amb les imatges histològiques de la tinció Hematoxilina-Eosina (HE), que és la que s'utilitza per detectar les zones tumorals.

L'objectiu principal ha estat entrenar una xarxa neuronal per predir, a partir de les WSI de la tinció HE, les zones amb més possibilitats de tenir cèl·lules tumorals. Concretament, el que s'ha fet és entrenar un model per tal de fer la segmentació de les imatges en 5 classes diferents: tumor, estroma, infiltració inflamatòria, necrosi i altres components. Així doncs, s'ha entrenat la xarxa amb una base de dades anotada anomenada *Breast Cancer Semantic Segmentation (BCSS)* i, un cop obtingut el millor model, s'ha fet inferència utilitzant les imatges de l'Institut Català de la Salut (ICS).

Altres objectius d'aquest treball han sigut desenvolupar aquestes eines de manera que puguin funcionar en un entorn real i que siguin eficients, tant en temps d'execució com en recursos computacionals. Un altre aspecte molt important és que aquesta xarxa ha de poder aprendre de noves dades anotades per tal d'adaptar-se a les imatges histològiques de l'ICS.

A partir del treball presentat, es pot dir que els objectius presentats anteriorment han estat assolits, ja que el resultat ha sigut un model entrenat amb el qual s'obtenen bons resultats al fer inferència amb les dades de l'ICS. Durant la realització d'aquest projecte, s'ha tingut molt en compte el programari i els recursos disponibles per part dels patòlegs per tal de que es pugui utilitzar el model resultant en un entorn el més real possible. A més, el model es pot ajustar a partir de noves dades anotades per tal d'aprendre d'aquestes i millorar els resultats. Finalment, comentar que la idea a l'hora d'aplicar el model obtingut és dividir la imatge histològica en *patches* de manera que la xarxa només predigui en les regions amb teixit i no en tota la *Whole Slide Image*.

Tot i això, hi ha aspectes que no s'han pogut fer degut a la falta de temps o de recursos. Així doncs, per tal de millorar els resultats d'aquest projecte, seria recomanable poder utilitzar anotacions de les imatges de l'Institut Català de la Salut per entrenar la xarxa, és a dir, per adaptar els seus pesos. Aquestes anotacions, però, poden obtenir-se a partir de la validació o correcció de les prediccions fetes per inferència amb el model obtingut. És a dir, que s'utilitzin les màscares resultants de la xarxa com a base de les futures anotacions dels patòlegs. D'aquesta manera, es redueix la càrrega de feina que suposa pels metges anotar aquestes imatges i es facilita la tasca d'obtenir un *dataset* de les imatges de l'ICS. Per tant, l'objectiu seria obtenir *feedback* per millorar l'anotació, reentrenar la xarxa a partir d'aquestes noves dades i millorar la qualitat de la predicció.

També seria interessant poder afegir noves classes per tal de fer prediccions més completes pels patòlegs. Respecte això, tal i com s'ha esmentat durant aquest treball, no s'ha pogut utilitzar la classe relativa als carcinomes ductals in situ, perquè no hi havia suficients anotacions. Així doncs, per aquest punt, també seria convenient poder tenir anotacions per parts dels patòlegs de l'ICS amb aquestes noves classes per entrenar la xarxa afegint aquest tipus d'elements. Una altra classe que seria

interessant de tenir anotada és la referent al teixit sa, ja que amb el model entrenat s'ha considerat l'estroma com la classe per defecte.

7. Bibliografía

- [1] A. Cruz-Roa *et al.*, “Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks,” in *Medical Imaging 2014: Digital Pathology*, Mar. 2014, vol. 9041, p. 904103, doi: 10.1117/12.2043872.
- [2] T. Araújo *et al.*, “Classification of breast cancer histology images using Convolutional Neural Networks,” *PLoS One*, vol. 12, no. 6, p. e0177544, Jun. 2017, doi: 10.1371/journal.pone.0177544.
- [3] J. GGecer, Baris & Aksoy, Selim & Mercan, Ezgi & Shapiro, Linda & Weaver, Donald & Elmore, “Detection and Classification of Cancer in Whole Slide Breast Histopathology Images Using Deep Convolutional Networks,” *Pattern Recognit.*, 2018, doi: 10.1016/j.patcog.2018.07.022.
- [4] M. Z. Alom *et al.*, “Advanced deep convolutional neural network approaches for digital pathology image analysis: A comprehensive evaluation with different use cases,” *arXiv*, vol. 2, 2019.
- [5] O. Iizuka, F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, and M. Tsuneki, “Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Dec. 2020, doi: 10.1038/s41598-020-58467-9.
- [6] A. P. Qu *et al.*, “Segmentation of Hematoxylin-Eosin stained breast cancer histopathological images based on pixel-wise SVM classifier,” *Sci. China Inf. Sci.*, vol. 58, no. 9, Sep. 2015, doi: 10.1007/s11432-014-5277-3.
- [7] Z. Guo *et al.*, “A Fast and Refined Cancer Regions Segmentation Framework in Whole-slide Breast Pathological Images,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.1038/s41598-018-37492-9.
- [8] G. Aresta *et al.*, “BACH: Grand challenge on breast cancer histology images,” *Med. Image Anal.*, vol. 56, pp. 122–139, Aug. 2019, doi: 10.1016/j.media.2019.05.010.
- [9] “¿Qué es el cáncer y cómo se desarrolla? - SEOM: Sociedad Española de Oncología Médica © 2019.” <https://seom.org/informacion-sobre-el-cancer/que-es-el-cancer-y-como-se-desarrolla>.
- [10] “¿Qué es el cáncer de mama?” https://www.breastcancer.org/es/sintomas/cancer_de_mama/que_es_cancer_mama.
- [11] “Cáncer de mama - Síntomas y causas - Mayo Clinic.” <https://www.mayoclinic.org/es-es/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470>.
- [12] N. M. Anderson and M. C. Simon, “The tumor microenvironment,” *Curr. Biol.*, vol. 30, no. 16, pp. R921–R925, Aug. 2020, doi: 10.1016/j.cub.2020.06.081.
- [13] “Invasión de los sistemas vascular o linfático.” https://www.breastcancer.org/es/sintomas/diagnostico/invasion_vascular_linfatico.
- [14] “Carcinoma ductal localizado - Síntomas y causas - Mayo Clinic.” <https://www.mayoclinic.org/es-es/diseases-conditions/dcis/symptoms-causes/syc-20371889>.

- [15] "Stroma | Pathology dictionary | MyPathologyReport.ca." <https://www.mypathologyreport.ca/stroma/>.
- [16] "Cáncer y Ejercicio. Capítulo 1. Hospital La Moraleja | Sanitas." <https://www.hospitallamoraleja.es/es/pacientes-cancer-deporte-capitulo1.php>.
- [17] "Necrosis Definition and Examples - Biology Online Dictionary." <https://www.biologyonline.com/dictionary/necrosis>.
- [18] "What is Staining? Definition, Objectives, Mechanism & Types - Biology Reader." <https://biologyreader.com/staining.html>.
- [19] G. Rolls and C. Sampias, "H&E Staining Overview: A Guide to Best Practices." Aug. 14, 2019.
- [20] P. A. Dorado, R. Celis, and E. Romero, "Color separation of H&E stained samples by linearly projecting the RGB representation onto a custom discriminant surface," in *12th International Symposium on Medical Information Processing and Analysis*, Jan. 2017, vol. 10160, p. 101600P, doi: 10.1117/12.2256966.
- [21] "Whole Slide Imaging | MBF Bioscience." <https://www.mbfbioscience.com/whole-slide-imaging>.
- [22] L. Pantanowitz, N. Farahani, and A. Parwani, "Whole slide imaging in pathology: advantages, limitations, and emerging perspectives," *Pathol. Lab. Med. Int.*, vol. 7, p. 23, Jun. 2015, doi: 10.2147/plmi.s59826.
- [23] W. Weng and X. Zhu, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, May 2021, doi: 10.1109/ACCESS.2021.3053408.
- [24] "U-Net. Introducing Symmetry in Segmentation | by Heet Sankesara | Towards Data Science." <https://towardsdatascience.com/u-net-b229b32b4a71>.
- [25] "How U-net works? | ArcGIS for Developers." <https://developers.arcgis.com/python/guide/how-unet-works/>.
- [26] "Softmax Activation Function — How It Actually Works | by Kiprono Elijah Koech | Towards Data Science." <https://towardsdatascience.com/softmax-activation-function-how-it-actually-works-d292d335bd78>.
- [27] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Nov. 2017, vol. 2017-Janua, pp. 5987–5995, doi: 10.1109/CVPR.2017.634.
- [28] "Review: ResNeXt — 1st Runner Up in ILSVRC 2016 (Image Classification) | by Sik-Ho Tsang | Towards Data Science." <https://towardsdatascience.com/review-resnext-1st-runner-up-of-ilsvrc-2016-image-classification-15d7f17b42ac>.
- [29] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2261–2269, Aug. 2016.
- [30] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Nov. 2017, vol. 2017-January, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.

- [31] “Review: Xception — With Depthwise Separable Convolution, Better Than Inception-v3 (Image Classification) | by Sik-Ho Tsang | Towards Data Science.” <https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568>.
- [32] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 10691–10700, May 2019.
- [33] “Google AI Blog: EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling.” <https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html>.
- [34] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Sep. 2015.
- [35] “VGG16 - Convolutional Network for Classification and Detection.” <https://neurohive.io/en/popular-networks/vgg16/>.
- [36] “IoU vs F1 score.” <https://tomkwok.com/posts/iou-vs-f1/>.
- [37] “F1/Dice-Score vs IoU - Cross Validated.” <https://stats.stackexchange.com/questions/273537/f1-dice-score-vs-iou>.
- [38] “Cross-Entropy Loss Function. A loss function used in most... | by Kiprono Elijah Koech | Towards Data Science.” <https://towardsdatascience.com/cross-entropy-loss-function-f38c4ec8643e>.
- [39] “A Gentle Introduction to Cross-Entropy for Machine Learning.” <https://machinelearningmastery.com/cross-entropy-for-machine-learning/>.
- [40] “Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures - Exsilio Blog.” <https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>.
- [41] “Precision, Recall, F1, Accuracy en clasificación - IArtificial.net.” <https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/>.
- [42] “Accuracy, Precision, Recall or F1? | by Koo Ping Shung | Towards Data Science.” <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>.
- [43] “Confusion Matrix for Machine Learning.” <https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/>.
- [44] “Breast Cancer Segmentation - Grand Challenge.” <https://bcsegmentation.grand-challenge.org/>.
- [45] “PathologyDataScience/CrowdsourcingDataset-Amgadetal2019. Computational and Integrative Pathology.” <https://github.com/PathologyDataScience/CrowdsourcingDataset-Amgadetal2019>.
- [46] M. Amgad *et al.*, “Structured crowdsourcing enables convolutional segmentation of histology images,” *Bioinformatics*, vol. 35, no. 18, pp. 3461–3467, Sep. 2019, doi: 10.1093/bioinformatics/btz083.
- [47] “Digital Slide Archive - A containerized web-based platform for the analysis, visualization, management and annotation of whole-slide digital pathology imaging data.” https://digitalslidearchive.github.io/digital_slide_archive/.

- [48] "Segmentation models with pretrained backbones. PyTorch." https://github.com/qubvel/segmentation_models.pytorch.
- [49] "What is image augmentation and how it can improve the performance of deep neural networks - Albumentations Documentation." https://albumentations.ai/docs/introduction/image_augmentation/.
- [50] "Cosine Annealing Explained | Papers With Code." <https://paperswithcode.com/method/cosine-annealing>.
- [51] "A Newbie's Guide to Stochastic Gradient Descent With Restarts | by Chi-Feng Wang | Towards Data Science." <https://towardsdatascience.com/https-medium-com-reina-wang-tw-stochastic-gradient-descent-with-restarts-5f511975163>.