

**UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO**



**DISEÑO DE UN PROCESO DE VALIDACIÓN DE
PATOLOGÍA DIGITAL EN CÁNCER DE MAMA**

JIMENA VICTORIA LOPEZ CAYO

**TESIS PARA OPTAR AL GRADO DE
MAGISTER EN INFORMÁTICA MÉDICA.**

Director de Tesis: Prof. Dr. Steffen Härtel

Co-Director de Tesis: Prof. Dra. Eugenia Díaz Guerra

2017

**UNIVERSIDAD DE CHILE
FACULTAD DE MEDICINA
ESCUELA DE POSTGRADO**

INFORME DE APROBACIÓN TESIS DE MAGISTER

Se informa a la Comisión de Grados Académicos de la Facultad de Medicina, que la Tesis de Magister presentada por la (el) candidata (o)

JIMENA VICTORIA LOPEZ CAYO

ha sido aprobada por la Comisión Informante de Tesis como requisito para optar al Grado de Magister en Informática Médica en el Examen de Defensa de Tesis rendido el día 28 de abril del 2017.

Prof. Dr. Steffen Härtel
Director de Tesis

COMISIÓN INFORMANTE DE TESIS

Prof. Dr. Mauricio Cerda

Prof. Dr. Ivan Gallegos

Prof. Dr. Patricia Gómez

Prof. Dr. Sergio Bozzo
Presidente Comisión

Dedicado a mis amados padres Irineo y Victoria.

A Nils y Teresa Spröhnle amada familia chilena.

A mis amados niños que recorren grandes distancias por acceso a una consulta
médica, mi motor, mi amor, mi inspiración.

Agradecimientos

Hubiera yo desmayado, si no creyese que veré la bondad de Jehová (Salmos 27:13). Doy gracias a Dios por su gracia, fidelidad y sustento en mi vida...tus planes son mucho mejores que los míos...Gracias.

Mi vida no alcanza para agradecer tanto amor y apoyo incondicional, a mis amados padres Irineo y Victoria quienes hicieron posible alcanzar esta meta.

A mi amada familia y familia de la fe por su constante apoyo y ánimo.

A la prestigiosa Universidad de Chile, al querido claustro nuclear del MIM y a los académicos que participaron en mi formación. A mis mentores, mi tutor Dr. Steffen Härtel y la Dra. Eugenia Díaz por su guía y enseñanza en el camino hacia la excelencia.

Al Servicio Alemán de Intercambio Académico (DAAD) por brindarme una beca y cubrir parte de mis estudios.

Agradezco especialmente a los integrantes y directivos del CPDAI y SCIAN por recibir al amigo extranjero y permitirme formar parte de su gran familia.

Mi reconocimiento a los patólogos Dr. Iván Gallegos, Dra. Paula Segura, Dr. Fernando Gabler, Dra. Yamile Corredoira, Dra. Marisol Goitia y Dra. Claudia Ramis por su participación al compartir sus conocimientos durante el desarrollo de la tesis.

Prof. Sandra de la Fuente...gracias por desafiarme a dar este paso, un sí o un no cambia rotundamente una vida.

Gracias a mis amigos cuya dosis de alegría, ocurrencias, tardes de almuerzo y café hicieron de este proceso una aventura, a Stefan Sigle, Sebastián Fernández, Patricio Araneda y Rodrigo Gálvez.

Como no mencionarte Juan Carlos Bonifacio Castro, mi admiración, cariño y respeto, gracias por tu ánimo, ayuda y las tardes de cuentos y relatos.

Una simple palabra que dice poco, pero expresa mucho...Gracias.

INDICE

RESUMEN	1
ABSTRACT.....	3
1. INTRODUCCION.....	5
1.1 PATOLOGÍA DIGITAL EN EL MUNDO Y EN CHILE	6
1.2 MOTIVACIÓN	7
1.3 CONTRIBUCIÓN / IMPACTO DE LA TESIS.....	9
1.4 HIPOTESIS	10
1.5 OBJETIVOS DE LA TESIS	10
1.6 ESTRUCTURA DE LA TESIS	11
2. MARCO TEÓRICO	12
2.1 CÁNCER DE MAMA	12
2.2 DIAGNÓSTICO DE BIOPSIAS EN CÁNCER DE MAMA.....	13
2.3 VARIABILIDAD EN LA EVALUACIÓN DE HER2 Y RECEPTORES DE HORMONAS.....	20
2.4 PATOLOGÍA DIGITAL.....	22
2.5 CLASIFICADORES DE APRENDIZAJE AUTOMÁTICO.....	29
2.6 MODELADOR DE PROCESOS.....	33
2.7 ESTADÍSTICOS DE CONCORDANCIA Y NORMALIDAD.	34
3. MATERIALES Y METODOS	38
3.1 ETAPA 1: MATERIALES Y DEFINICIÓN DEL FLUJO DE TRABAJO	38
3.2 ETAPA 2: GENERACIÓN DE GOLD STANDARD POR PATÓLOGOS.....	42
3.3 ETAPA 3: ANÁLISIS DE IMAGEN	45
3.4 ETAPA 4: ANÁLISIS ESTADÍSTICO DE CONCORDANCIA PATÓLOGOS VS ALGORITMOS	57

4. RESULTADOS	58
4.1 IMPLEMENTACIÓN	58
4.2 EVALUACIÓN VISUAL DE PLACAS VIRTUALES – GOLD STANDARD....	59
4.3 CUANTIFICACIÓN IHQ EN HER2 Y ANÁLISIS ESTADISTICO DE CONCORDANCIA.....	66
4.4 CUANTIFICACIÓN IHQ EN RE Y ANÁLISIS ESTADISTICO DE CONFIABILIDAD.....	75
5. DISCUSIÓN	87
5.1 FLUJO DE TRABAJO DEL PROCESO DE VALIDACIÓN: CONECTANDO PATÓLOGOS Y ORGANIZACIONES.....	87
5.2 PLACAS VIRTUALES Y GENERACIÓN DE UN GOLD STANDARD	88
5.3 CUANTIFICACIÓN AUTOMATIZADA Y VARIABILIDAD EN HER2	88
5.4 CUANTIFICACIÓN AUTOMATIZADA Y VARIABILIDAD EN RECEPTORES HORMONALES.....	91
6. RESUMEN & CONCLUSIÓN	95
7. BIBLIOGRAFÍA	97
8. ANEXOS	107

INDICE DE TABLAS

Tabla 4.1: Comparación de resultados de la clasificación general realizada por patólogos versus CART.	68
Tabla 4.2: Concordancia observada en ROIs distribuidas por patólogo.....	69
Tabla 4.3: Concordancia observada patólogos versus CART en clasificaciones generales.....	69
Tabla 4.4: Comparación de resultados de la clasificación general realizada por patólogos versus SMO.	72
Tabla 4.5: Concordancia observada en ROIs distribuidas por patólogo.....	73
Tabla 4.6: Resultados de la prueba de concordancia Cohen´s Kappa entre tres patólogos (P1-3).....	73
Tabla 4.7: Concordancia patólogos versus SMO	74
Tabla 4.8: Coeficientes RE de 100% positividad de 10 ROIs (cc_1-10) y promedios (cp ₁₀₀) en 10 placas.....	76
Tabla 4.9: Comparación de la evaluación del porcentaje de núcleos positivos para RE realizada por tres patólogos (P1-3) versus el algoritmo RG (np _{pos}).....	78
Tabla 4.10: Resultados del test de concordancia K-alpha entre tres patólogos (P1-3).	79
Tabla 4.11: Resultados de concordancia K-alpha de tres patólogos (P1-3) versus el resultado del algoritmo RG	79
Tabla 4.12: Resultados de concordancia K-alpha de tres patólogos (P1-3) versus el resultado del algoritmo RG	79
Tabla 4.13: Coeficientes RP de 100% positividad de 10 ROIs (cc_1-10) y promedios (cp ₁₀₀).	81
Tabla 4.14: Tabla de evaluación de tres patólogos (P1-3) y el resultado del algoritmo RG para RP.....	83

Tabla 4.15: Resultados de concordancia K-alpha para RP entre patólogos de sin RP_0005/11 (*).....	84
Tabla 4.16: Resultados de concordancia K-alpha para RP entre tres patólogos versus el resultado del algoritmo RG de los 10 casos; sin RP_0005/11 (*).....	85
Tabla 4.17: Resultados de concordancia K-alpha para RP de tres patólogos versus el resultado del algoritmo RG en 82 ROIs analizados; sin RP_0001/5/11 (*) N = 59 ROIs.	85

INDICE DE FIGURAS

Figura 2.1: Células del conducto lácteo.	12
Figura 2.2: Biopsia.	13
Figura 2.3: Pruebas de IHQ y FISH para detectar células cancerígenas por la presencia de HER2.	15
Figura 2.4: Algoritmo para la evaluación receptor 2 del factor de crecimiento epidérmico humano (HER2)..	16
Figura 2.5: Ejemplo de la evaluación IHQ de HER2 según clasificación ASCO/CAP 0, 1+, 2+ y 3+..	17
Figura 2.6: determinación del porcentaje de positividad nuclear en receptores hormonales.....	19
Figura 2.7: Ejemplo de la evaluación IHQ de RE según recomendación ASCO/CAP	19
Figura 2.8:Resultados del control de calidad externo realizado por NordiQC..	22
Figura 2.9: Transformación de un flujo de trabajo manual a un flujo de trabajo digital	23
Figura 2.10: Plataforma de trabajo del software NDP.analyze de análisis de imagen de Visiopharm..	26
Figura 2.11: ROI de una biopsia mamaria analizada por el algoritmo HER2 configurado bajo dos valores distintos de sensibilidad	28
Figura 2.12: Selección de núcleos de entrenamiento en una muestra para RE.....	29
Figura 2.13: Algoritmo de árboles de decisión en máquinas de aprendizaje..	31
Figura 2.14: Ilustración simplificada de una clasificación lineal mediante SVM.	32
Figura 2.15: Modelado del proceso de triage en un centro médico.....	33
Figura 3.1: Etapas del proceso de validación de PD en CM..	38

Figura 3.2: Proceso manual de exportación de ROIs desde la plataforma NDP.serve a la plataforma de Visiopharm.....	41
Figura 3.3: Proceso de evaluación de placas virtuales para los casos de HER2, RE y RP	43
Figura 3.4: Ejemplo de ROIs seleccionadas y evaluadas por los patólogos	43
Figura 3.5: Matriz de localización de ROIs para establecer coincidencias entre patólogos.....	44
Figura 3.6: Identificación y etiquetado de núcleos positivos y negativos para RE.....	45
Figura 3.7: Configuración del parámetro Sensibilidad para la obtención de los valores de conectividad para HER2.....	47
Figura 3.8: Coeficientes de fluctuación de grupo 3+ respecto a 1+ en función de la sensibilidad.....	48
Figura 3.9: Relación sensibilidad – conectividad.....	49
Figura 3.10: Proceso de aprendizaje del clasificador (algoritmo) a partir de datos del conjunto de entrenamiento.	50
Figura 3.11: Etiquetado de núcleos para entrenamiento del algoritmo de cuantificación.....	52
Figura 3.12: Detección de núcleos mediante el algoritmo Visiopharm Nuclei App... ..	52
Figura 3.13: Secuencia de detección específica de núcleos positivos mediante el algoritmo RG.	54
Figura 3.14: Imagen representativa de un control positivo en una placa.	55
Figura 4.1: Flujo de trabajo seguido por los patólogos.....	59
Figura 4.2: Variabilidad en el tamaño de las ROIs dibujada por los patólogos.....	60
Figura 4.3: Numero de coincidencia entre la localización de ROIs.	61
Figura 4.4: Variabilidad en la clasificación de áreas de 3 y 2 hits..	62

Figura 4.5: Ejemplo de variabilidad en la clasificación de ROIs en áreas de 3 y 2 hits	63
Figura 4.6: Concordancia y variabilidad inter-observador..	64
Figura 4.7: Traducción de valores de conectividad a clasificación ASCO/CAP mediante metodología CART.	67
Figura 4.8: Traducción de valores de conectividad a clasificación ASCO/CAP mediante metodología SMO.....	71
Figura 4.9: Variabilidad de coeficientes RE entre ROIs y placas (datos de Tabla 4.8)	77
Figura 4.10: Variabilidad de coeficientes RP entre ROIs y placas (datos de Tabla 4.13).	82

RESUMEN

El advenimiento de la Patología Digital (PD) y la creación de placas virtuales ofrecen nuevas posibilidades para el procesamiento, análisis e interpretación de imágenes que se suman al creciente interés en la cuantificación de células y/o biomarcadores para dar apoyo a la decisión diagnóstica, pronóstica y terapéutica sobre todo en patologías como el cáncer [1-3]. En este sentido, la Asociación Americana de Oncología Clínica (ASCO) y el Colegio Americano de Patólogos (CAP) han mencionado el uso del análisis de imágenes, como método deseable para cuantificar marcadores inmunohistoquímicos en cáncer de mama (CM), tales como el receptor 2 del factor de crecimiento epidérmico humano (HER2) y receptores hormonales [4-6]. Actualmente, en países desarrollados (norteamericanos, europeos y asiáticos) se han realizado estudios que validan este tipo de procedimientos [7-12], sin embargo, no existen antecedentes de que estos hayan sido realizados en Chile, ya que esta tecnología está recientemente presente en el territorio nacional. Hoy en día el Centro de Patología Digital Asistido por Internet (CPDAI) cuenta con estos recursos abriendo una oportunidad de uso y aplicación de esta tecnología como apoyo al diagnóstico clínico en la cual se hace imperiosa su validación.

En el presente proyecto se diseñó e implementó un proceso de validación de DP y su aplicación específica en CM en un contexto nacional. Se estandarizaron algoritmos pre-definidos del *software* comercial Visiopharm para la cuantificación de marcación inmunohistoquímica (IHQ) de HER2 y receptores hormonales. Además, permitió proponer una nueva metodología de cuantificación nuclear usando las herramientas existentes en el software, que se denominó *Region Growing (RG)*.

El flujo de trabajo del proceso de validación fue representado en la notación estándar de modelo de negocios (*Business Process Model and Notation, BPMN*). La estandarización y validación de los algoritmos de cuantificación incluyó 46 biopsias de tejido mamario procesadas en IHQ, de las cuales 26 corresponden a HER2, 10 a receptores de estrógeno (RE) y 10 a receptores de progesterona (RP). Los algoritmos fueron ajustados y entrenados a la evaluación diagnóstica de 3 patólogos

con experiencia (*gold standard*) en un conjunto de entrenamiento y verificadas en un conjunto de validación. En HER2 la concordancia observada entre la evaluación de los patólogos versus la cuantificación mediante algoritmos fue de $k = 0.51$ (71.4%) que corresponde a un nivel de concordancia moderado similar a la observada entre los patólogos que varía entre $k = 0.27$ a 0.51 . Los resultados obtenidos mediante la cuantificación con el algoritmo *RG* para receptores hormonales fueron comparados según niveles de confiabilidad de *Krippendorff's alpha*. Para los 10 casos de RE se obtuvo un $k\text{-alpha} = 0,99$ que corresponde a un nivel altamente confiable, en tanto que para RP el valor de $k\text{-alpha} = 0.2$ con nivel no confiable.

La baja concordancia encontrada entre la cuantificación mediante algoritmo versus la evaluación de patólogos se puede explicar con factores como i) variabilidad inter-observador, ii) heterogeneidad intrínseca del tejido mamario y iii) variabilidad en los procesos de preparación del tejido y tinción IHQ. Los dos últimos factores dificultan tanto la evaluación realizada por los patólogos como la cuantificación mediante algoritmo.

A través de este primer estudio se ha podido establecer, implementar y estandarizar un conjunto de procedimientos y protocolos para la implementación y uso de las herramientas de la PD en la realidad nacional. Los algoritmos de análisis de imágenes presentan valores de concordancia y confiabilidad equivalentes a la evaluación realizada por los patólogos. Este método puede ser optimizado en términos de concordancia, mejorando variables previas a la cuantificación, como, por ejemplo, el procesamiento del tejido y la calidad de la tinción IHQ, a través de la estandarización de dichos procesos y/o contar con controles de calidad en IHQ externos. Los avances actuales junto con la mejora de la tecnología, sin duda abrirán el camino para una mayor adopción a mediano y largo plazo, permitiendo a la comunidad de patólogos en su conjunto aprovechar el verdadero potencial de la PD para el tratamiento y cuidado del paciente.

ABSTRACT

The advent of Digital Pathology (PD) and the use of Whole Slide Imaging (WSI) offer new possibilities for the processing, analysis and interpretation of images that add to the growing interest in the quantification of cells and / or biomarkers to support the Diagnostic, prognostic and therapeutic decision, especially in pathologies such as cancer [1-3]. In this regard, the American Society of Clinical Oncology (ASCO) and the American College of Pathologists (CAP) have mentioned the use of imaging as a desirable method for quantifying immunohistochemical markers in breast cancer (BC), such as the receptor 2 of human epidermal growth factor (HER2) and hormone receptors [4-6].

Currently, studies have been carried out in developed countries (Denmark, United States, Canada, etc) that validate this type of procedures [7-9]. However, there is no evidence that these have been carried out in Chile, since this technology is recently present on the national territory. Today the Center for Digital Pathology Assisted by Internet (CPDAI) has these resources opening an opportunity to use and apply this technology to support the clinical diagnosis in which it becomes imperative to validate.

In this project, a validation process of DP was designed and implemented, and its specific application in BC in a local context was presented. The default algorithms of commercial software Visiopharm for the quantification of immunohistochemical staining (IHQ) of HER2 and hormone receptors were standardized. Although, a new methodology of nuclear quantification using the available tools in the software was proposed. This new methodology was called Region Growing (RG).

The workflow of the validation process was represented in the standardized graphical representation called standard business process model and notation (BPMN). The standardization and validation of the algorithms included 46 breast tissue biopsies processed in IHC, of which 26 corresponded to HER2, 10 to estrogen receptors (ER) and 10 to progesterone receptors (PgR). The algorithms were to match the diagnostic evaluation of 3 experienced pathologists (gold standard) in a

training set and verified in a validation set in HER2 the agreement observed between the evaluation of pathologists versus the quantification using algorithms was by Cohen's $k = 0.51$ (71.4%) corresponding to a moderate concordance level similar to that observed among pathologists ranging from $k = 0.27$ to 0.51 . The results obtained for hormone receptors were compared according to levels of reliability of Krippendorff's alpha. For the 10 cases of RE, a $k\text{-alpha} = 0.99$ corresponding to a highly reliable level was obtained, while for RP the value of $k\text{-alpha} = 0.2$ with an unreliable level.

The low agreement found between algorithm quantification versus pathologist evaluation could be explained by factors such as (i) interobserver variability, (ii) intrinsic heterogeneity of breast tissue and (iii) variability in tissue preparation and IHC staining. The last two factors make it difficult for both pathologists to evaluate and algorithm quantification.

Through this first study it has been possible to establish, implement and standardize a set of procedures and protocols for the use and application of PD tools in the local reality. Image analysis algorithms have values of concordance and reliability equivalent to the evaluation performed by pathologists. This method can be optimized in terms of agreement, improving pre-quantification variables, such as tissue processing and IHQ staining quality, through the standardization of these processes. Current advances in Chile coupled with improved technology will undoubtedly pave the way for greater medium- and long-term adoption, allowing the community of pathologists as a whole to tap into the true potential of PD for patient care.

1.INTRODUCCION

La Anatomía Patológica (AP) es el estudio de la naturaleza de las enfermedades, sus causas, desarrollo y consecuencias. Se basa en el examen macroscópico y microscópico de tejidos o biopsias procesadas mediante técnicas histoquímicas e inmunohistoquímicas (IHQ), desde las cuales se extrae información relevante para el diagnóstico y pronóstico de la enfermedad, así como el tratamiento de los pacientes [13-15]. La práctica tradicional de la AP requiere la evaluación y análisis visual de placas histopatológicas a través de un microscopio convencional [14-16]. Esta metodología de trabajo puede llegar a inducir cierto grado de subjetividad en el diagnóstico clínico, en particular en casos en que se requiere evaluar marcadores biológicos específicos a nivel celular o molecular [11, 12]. Se enfrenta además con el hecho que, de ser necesaria una segunda opinión o interconsulta, se requiere el envío de placas a otro especialista, lo que agrega tiempo adicional a la entrega de resultados [16].

En patologías en las que el análisis de biopsias determina el diagnóstico médico y ulterior tratamiento, como por ejemplo el cáncer, la falta de mediciones precisas y reproducibles puede llegar a constituir un importante obstáculo para el desarrollo de terapias [11, 12, 17, 18] . Esto demanda por parte de los patólogos una mayor precisión y tiempo dedicado al análisis [16-18]. Con el propósito de hacer más precisa la evaluación e interpretación de biopsias y marcadores biológicos, instituciones como la ASCO/CAP han desarrollado protocolos y guías de recomendación que son utilizadas como referentes a nivel mundial. Sin embargo, pese a tales esfuerzos, diversos estudios clínicos ponen en evidencia la existencia de un importante grado de variabilidad en el análisis de biopsias, tanto intra- como inter-observador [11, 12, 19, 20].

En la última década, los avances tecnológicos en software y hardware han hecho posible la digitalización de placas histológicas y la creación de placas virtuales, a través de microscopios robotizados, conocidos como escáner de tejido de funcionamiento automatizado [1, 21, 22]. Las placas virtuales así generadas son

archivos de datos de imágenes de toda la placa digitalizada a alta resolución [1, 23]. Estas son visualizadas mediante plataformas informáticas que emulan las funcionalidades de un microscopio [1], potenciándose con las tecnologías de información a través del internet y accedidas mediante el uso e intercambio de datos e imágenes on-line a través de computadores o dispositivos móviles [1, 21, 22]. Estos importantes desarrollos tecnológicos ofrecen grandes ventajas sobre la microscopía tradicional, al permitir en forma muy expedita facilidad de acceso, interconsulta, segunda opinión, así como múltiples aplicaciones y herramientas de utilidad en el diagnóstico clínico, investigación y educación [1, 21-24].

Como consecuencia de lo anterior, la AP, así como la mayoría de las especialidades médicas, ha experimentado una transformación que va desde un flujo de trabajo manual a un flujo de trabajo digital, lo que se conoce como Patología Digital (PD) [1, 16, 21, 22]. Este cambio de paradigma desde lo analógico hacia lo digital, viene acoplado con nuevas herramientas tecnológico-matemáticas que permiten dar un salto importante en el procesamiento de señales y patrones biológicos, incrementando no solo el nivel de precisión del análisis, sino que, además abriendo la posibilidad de extraer nueva información, invisible al ojo humano [1, 16, 21].

Con el advenimiento y masificación del uso de placas virtuales, se ha suscitado un gran interés en el desarrollo y uso de algoritmos de análisis de imagen que permiten cuantificar diversos marcadores biológicos. Los beneficios aportados por tales algoritmos incluyen mejoras en exactitud y precisión en la detección, clasificación y medición de patrones morfométricos [1, 11, 12, 16, 17]. En este último aspecto, estos análisis actualmente están siendo aplicados principalmente en la cuantificación de la expresión de marcadores tumorales en cáncer [1, 11, 12, 25].

1.1 PATOLOGÍA DIGITAL EN EL MUNDO Y EN CHILE

Actualmente, la incorporación y utilización de PD está notablemente avanzada en Norteamérica (Estados Unidos y Canadá), países europeos tales como Alemania [22], Dinamarca [4], España [26] y en Japón [27]. En todos estos países existe una

amplia experiencia en la incorporación de PD con propósitos diagnósticos, de investigación, académicos [22] y epidemiológicos [25].

En Chile y Latinoamérica no había presencia de PD hasta el 2014. En dicho año surgió el primer Centro de Patología Digital Asistido por Internet (CPDAI) en el Laboratorio SCIAN-Lab de la Facultad de Medicina de la Universidad de Chile¹, desarrollado a través del Proyecto FONDEF D11I1096. CPDAI cuenta con equipamiento e infraestructura de vanguardia y es pionero en Chile y Latinoamérica, ofreciendo los servicios de digitalización de placas histológicas, análisis cuantitativo de patrones biológicos, almacenamiento (*storage service*), cursos online de MV y PD y destacando por su tecnología sofisticada y capacidad de funcionamiento.

En este contexto se desarrolla la presente tesis de Magister, enfocada a la estandarización y validación de las herramientas de PD y su aplicación específica en la cuantificación de los principales marcadores tumorales en cáncer de mama (CM). Esto se abordó mediante el diseño e implementación de un proceso de estandarización y validación que involucró diversas etapas de ajuste y configuración, requeridas para validar estas herramientas tecnológicas de manera que se constituyan en un efectivo apoyo al diagnóstico clínico en un contexto local.

1.2 MOTIVACIÓN

La Organización Mundial de la Salud (OMS) considera el CM como uno de los principales problemas de salud pública en el mundo², siendo una patología frecuente en mujeres, tanto en países desarrollados como en vías de desarrollo. Estimaciones realizadas por GLOBOCAN³ el 2012 indican que se registraron alrededor de 1.67 millones de nuevos casos de CM a nivel mundial, alcanzando un total de 522.000 fallecimientos por esta causa [28]. En América Latina y el Caribe, la misma fuente reportó que 408.000 mujeres fueron diagnosticadas con CM y más de 92.000 murieron en la región a causa de esta patología. Se espera que para el año 2030

¹ www.microscopiavirtual.cl. Accedido el: 19-10-2016

² <http://www.wcrf.org/int/cancer-facts-figures/worldwide-data>. Accedido el: 19-10-2016

³ GLOBOCAN: proyecto de salud llevado a cabo por la *International Agency for Research on Cancer (IARC)* que tiene como objetivo proporcionar las estimaciones actuales de incidencia y mortalidad de los principales tipos de cáncer, a nivel nacional, para todos los países del mundo.

esta cifra aumente en un 46%¹, lo que se traduciría en que anualmente fallecerían 142.100 mujeres y se producirían 596.000 casos nuevos, de mantenerse las tendencias actuales. La proyección de este incremento en número de nuevos casos y muertes será casi el doble en América Latina y el Caribe en comparación con América del Norte [28].

En Chile, el CM es una de las principales causas de muerte en mujeres². En el año 2012 se produjeron 1.367 fallecimientos, con una tasa estandarizada por edad de 15.1 por 100.000 mujeres [29]. El primer informe de registros poblacionales de cáncer, presentado por el Ministerio de Salud publicado el 2012, reporta una incidencia que varía entre 29.6 y 32.3 casos nuevos por 100.000 mujeres [29], afectando con mayor frecuencia a mujeres mayores de 15 años [29-31]. La evolución de la mortalidad en CM muestra un aumento constante, del orden del 20% superando 1.300 muertes anuales [29, 30]. Esta patología es una importante carga para la salud, tanto para los pacientes como para los sistemas de salud pública, presentando consecuencias sociales y psicológicas significativas [28]. Por tanto, una detección temprana unida a un diagnóstico y tratamiento adecuado, constituyen las mejores herramientas disponibles en la actualidad para disminuir las muertes por CM³ [28-31].

Tomando en consideración la dotación de especialistas dedicados a esta patología, los datos proporcionados por la Superintendencia de Salud indican que en el territorio nacional solo se encuentra registrado un total de 270 patólogos⁴, para atender una población de género femenino de 8.000.000⁵. Bajo esta perspectiva, el principal objetivo de esta tesis es proveer una ayuda a los especialistas en el proceso de diagnóstico, a través del uso de herramientas tecnológicas como el *Digital Slide Scanner* (microscopio robotizado) de placas histológicas y la utilización de algoritmos matemáticos que permiten realizar cuantificación automatizada de patrones biológicos.

¹ http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&gid=17926&Itemid=270

² www.deis.cl, Accedido el: 25-05-2015

³ http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&gid=17926&Itemid=270. Accedido el: 25-05-2015.

⁴ www.supersalud.gob.cl/568/w3-propertyvalue-3929.html, Accedido el: 25-05-2015

⁵ www.ine.cl/canales/menu/.../pdf/.../estadisticas_demograficas_2012.pdf, Accedido el: 25-05-2015

Recientemente, la compañía danesa Visiopharm ha colocado en el mercado un innovador *software* de análisis de imagen, el cual ya ha sido instalado en 7 hospitales europeos (Hospitales Universitarios de Aarhus y Aalborg, entre otros) para su uso clínico y de diagnóstico¹. En esta línea, CPDAI adquirió un paquete de *software* de esta compañía y ha procedido a estandarizar y optimizar su utilización en el entorno nacional, siendo esta una de las más importantes contribuciones que entrega este trabajo de tesis.

1.3 CONTRIBUCIÓN / IMPACTO DE LA TESIS

Los resultados obtenidos en esta tesis tendrán un impacto en el ámbito social y de salud pública. Desde el punto de vista clínico, se entrega una herramienta de apoyo al diagnóstico, y se genera una metodología de aplicación tanto para los especialistas en AP, como para el centro que presta el servicio (CPDAI).

Desde el punto de vista de la investigación científica, la plataforma del software de análisis de imágenes de Visiopharm puede ser usada potencialmente para configurar y/o crear una variedad de algoritmos para cuantificación de otros marcadores tumorales, en este contexto, la metodología de interacción llevada a cabo en esta tesis puede ser seguida para estandarizar estos nuevos algoritmos, como por ejemplo la cuantificación de HER2 en cáncer gástrico donde umbrales de corte pueden ser definidos a través de modificaciones o adaptaciones específicas a la metodología propuesta. Así mismo para poder obtener un *gold standard* es posible seguir la metodología de selección de ROIs, brindando acceso *online* a los patólogos, despejando así limitantes antes condicionadas por la presencia in situ del especialista. Además, se ha abierto una nueva línea de investigación en lo referente a la calidad del procesamiento de muestras histopatológicas.

Finalmente, la metodología desarrollada en esta tesis puede ser usada como referente para la realización de estudios de variabilidad intra e inter-observador, tales como pruebas de competencia para especialistas, a través del uso de placas virtuales.

¹ www.visiopharm.com/blog/tag/image-analysis/page/2. Accedido el: 20-08-2015

1.4 HIPOTESIS

Los parámetros requeridos para estandarizar el uso de algoritmos del software Visiopharm en la cuantificación automatizada de marcadores tumorales en cáncer de mama pueden ser obtenidos a través del diseño e implementación de un proceso de validación de Patología Digital.

1.5 OBJETIVOS DE LA TESIS

El objetivo general y los objetivos específicos de este trabajo son:

Objetivo general:

Diseñar e implementar un flujo de trabajo de Patología Digital que permita estandarizar y validar algoritmos de Visiopharm para cuantificar los marcadores inmunohistoquímicos HER2, RE y RP en biopsias de cáncer de mama en Chile.

Objetivos específicos:

1. Diseñar y elaborar el diagrama del flujo de trabajo del proceso de validación.
2. Obtener la evaluación de HER2, realizada por patólogos en biopsias de tejido mamario procesadas con IHQ (*gold standard*).
3. Obtener la evaluación de positividad nuclear de RE y RP, realizada por patólogos en biopsias de tejido mamario procesadas en IHQ (*gold standard*).
4. Establecer coeficientes y parámetros para configurar los algoritmos Visiopharm.
5. Cuantificar mediante los algoritmos la expresión IHQ de los marcadores utilizados.
6. Comparar los resultados de la cuantificación mediante algoritmos versus patólogos.

1.6 ESTRUCTURA DE LA TESIS

La tesis está estructurada en 5 Capítulos. El Capítulo 1 consiste en una breve introducción a la problemática abordada.

El Capítulo 2 presenta el estado del arte, fundamento teórico y conceptos fundamentales en los que se sustenta este trabajo, como son el cáncer de mama, el diagnóstico, componentes biológicos, su evaluación y clasificación, relevancia del análisis y problemas frecuentes. Se definen además conceptos técnicos tales como placas virtuales, visualizadores, análisis de imagen, clasificadores de máquinas de aprendizaje y la descripción del software y algoritmos utilizados en la tesis.

El Capítulo 3 describe el diseño y construcción de la metodología propuesta y su posterior implementación. Se describen las tareas y actividades realizadas en CPDAI y la participación de los patólogos en la generación del *ggold standard*, configuración de los algoritmos y la clasificación automatizada.

El Capítulo 4 presenta los resultados obtenidos de la metodología propuesta, tanto para la obtención del *gold standard*, cuantificación de HER2 y RE/RP mediante análisis de imagen en el contexto de la PD y el análisis estadístico de la concordancia entre la evaluación del patólogo versus la obtenida mediante el algoritmo.

En el Capítulo 5 se discuten los principales hallazgos y dificultades encontradas, así como posibles mejoras a realizar.

Finalmente, el Capítulo 6 terminando con nuestras conclusiones y las proyecciones que surgen de este trabajo.

2.MARCO TEÓRICO

2.1 CÁNCER DE MAMA

El CM es una enfermedad en la que algunas células del tejido mamario se vuelven anormales, multiplicándose de manera incontrolable. Se manifiesta de manera localizada hasta llegar a formar un tumor y puede diseminarse a otros tejidos circundantes, evento conocido como metástasis [29, 32-35]. A pesar que el CM es mucho más común en mujeres, este tipo de cáncer también se puede desarrollar en hombres [31]. La mayoría de los CM comienzan en las células de los conductos lácteos (cánceres ductales; Ver esquema en Figura 2.1 o se presentan en las células de la porción secretora de las glándulas que producen la leche materna (cánceres lobulares), aunque también pueden originarse en los tejidos estromales, que incluyen los tejidos conectivos adiposos y fibrosos de la mama [34, 35]. La anomalía surge como consecuencia de la acumulación de alteraciones genéticas y epigenéticas, las que provocan cambios a nivel molecular, bioquímico y celular [29, 32].

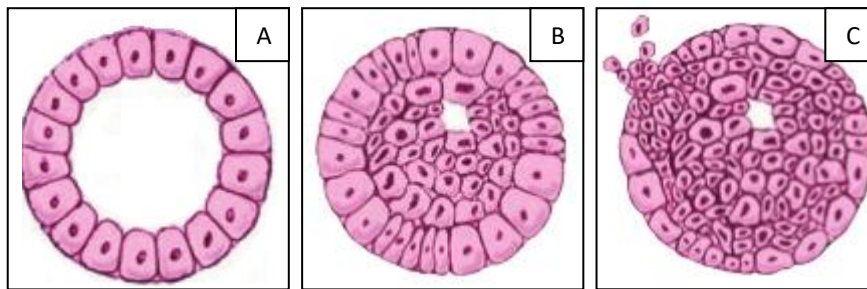


Figura 2.1: Células del conducto lácteo: A) Células epiteliales normales formando la pared del conducto; B) Células cancerígenas proliferativas no invasivas, permanecen dentro de los conductos lácteos; C) Células cancerígenas invasivas, abandonan el ducto y migran hacia los tejidos normales circundantes. Fuente: www.breastcancer.org/, "Tu guía para el informe patológico del cáncer de mama". Accedido el: 21-11-2016.

La información biológica más específica del CM es obtenida mediante una evaluación histopatológica a través de un examen de biopsia de mama.

2.2 DIAGNÓSTICO DE BIOPSIAS EN CÁNCER DE MAMA

La biopsia consiste en la extracción de un trozo del tejido de la zona anómala (lesión neoplásica) con el propósito de constatar la presencia de células cancerígenas [29] (Figura 2.2 A)¹.

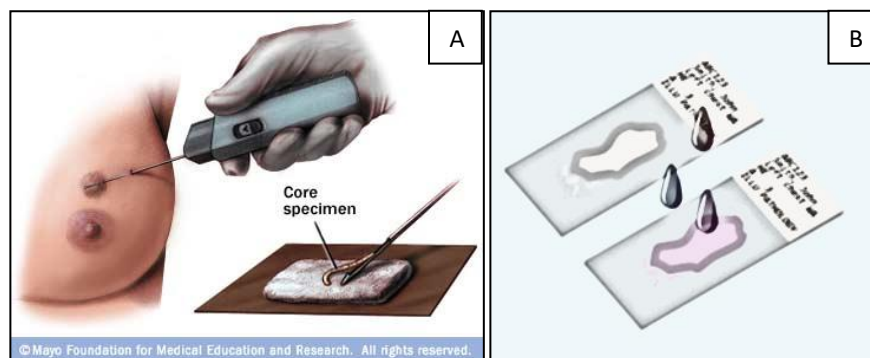


Figura 2.2: Biopsia: A) Extracción de tejido neoplásico en la mama, mediante un proceso de punción de aguja denominado *biopsia core*; B) Representación de biopsias procesadas mediante técnicas HQ e IHQ que permiten revelar la presencia o ausencia de proteínas específicas en células cancerígenas en muestras de tejido neoplásico mamario. Fuente: www.mayoclinic.org/tests-procedures/needle-biopsy/multimedia/core-needle-biopsy/img-20008370. Accedido el: 21-11-2016.

Las células anormales o cancerígenas son detectadas mediante el uso de proteínas o sustancias conocidas como marcadores tumorales o biomarcadores, producidas en grandes concentraciones por estas células u otras células del cuerpo como consecuencia del cáncer [29, 32]. Estos marcadores tumorales son ampliamente utilizados tanto para prevención y detección, como para determinar tratamiento y seguimiento a pacientes [35, 36]. En CM existen diversos marcadores tumorales, pero la información pronóstico más específica se obtiene mediante dos marcadores principales: i) los receptores de hormonas (receptor de estrógeno, RE y receptor de progesterona, RP) y ii) el receptor 2 del factor de crecimiento epidérmico humano (*human epidermal growth factor receptor 2*, HER2), cuya pesquisa proporciona información sobre el nivel que ha alcanzado la patología, que fármaco específico se puede prescribir al paciente y la evolución de la patología al tratamiento [12, 15, 19, 36, 37].

¹ <http://www.i-d.cl/biopsiacore.html>. Accedido el: 21-11-2016

Debido a la relevancia de los marcadores mencionados, organizaciones como la Sociedad Americana de Oncología Clínica (ASCO) y el Colegio Americano de Patólogos (CAP), han desarrollado guías y recomendaciones conocidas y usadas como referentes a nivel mundial para evaluar el estado de estos marcadores tumorales de manera sistemática [11, 12, 19]. Sin embargo, pese a los esfuerzos realizados, la literatura en el campo indica la existencia de un importante grado de variabilidad en la evaluación de estos marcadores [11, 12, 20, 38].

2.2.1 Consideraciones acerca de la evaluación de HER2

Es ampliamente sabido que el gen HER2/neu puede influir en el desarrollo del CM [39]. Este gen produce proteínas HER2 que son receptores de membrana presentes en células mamarias normales, cuyo rol es controlar su división y crecimiento en condiciones normales [38, 39]. En CM este gen no funciona correctamente y produce muchas copias de sí mismo (proceso conocido como amplificación del gen HER2/neu), lo que conlleva a una sobreexpresión de la proteína HER2 [19, 38-40], haciendo que las células mamarias crezcan y se dividan de manera incontrolable. Tanto la amplificación del gen como la sobreexpresión de la proteína HER2 están asociados a un tipo más agresivo de CM y por ende a un mal pronóstico para las pacientes [19, 38-40]. La sobreexpresión del gen HER2 se encuentra presente en aproximadamente un 15% a 20% de los CM [19, 40] y constituye un importante factor pronóstico y predictivo.

Clínicamente, el estado de HER2 es evaluado principalmente para determinar la elegibilidad del paciente a beneficiarse con una terapia específica cuyo fármaco más conocido es el Trastuzumab [38, 40]. Para la detección tanto a nivel genético como a nivel de proteína (Figura 2.3) existen diferentes métodos y técnicas, sin embargo, la técnica más ampliamente usada, tanto por factores de costo como de acceso, es la técnica de inmunohistoquímica (IHQ) para detección de la proteína.

2.2.2 Evaluación de HER2 en biopsias procesadas con IHQ

La evaluación de HER2 mediante IHQ utiliza criterios de completitud de membrana e intensidad de la inmunotinción. El análisis IHQ indica si la existencia de una elevada cantidad de la proteína HER2 (sobrexpresión) presente en la membrana celular de las células cancerígenas o si no la hay (negativo, sin sobreexpresión de la proteína). Los resultados del análisis de IHQ se clasifican de acuerdo a la escala: 0 (negativo), 1+ (negativo), 2+ (equivoco), o 3+ (positivo). Si el resultado de la evaluación es equivoco (2+), se requiere de una prueba adicional a nivel genético, la que se realiza mediante *hibridación fluorescente in situ (FISH)* [19]. En la Figura 2.3 se ilustran de manera comparativa ambas técnicas de análisis.

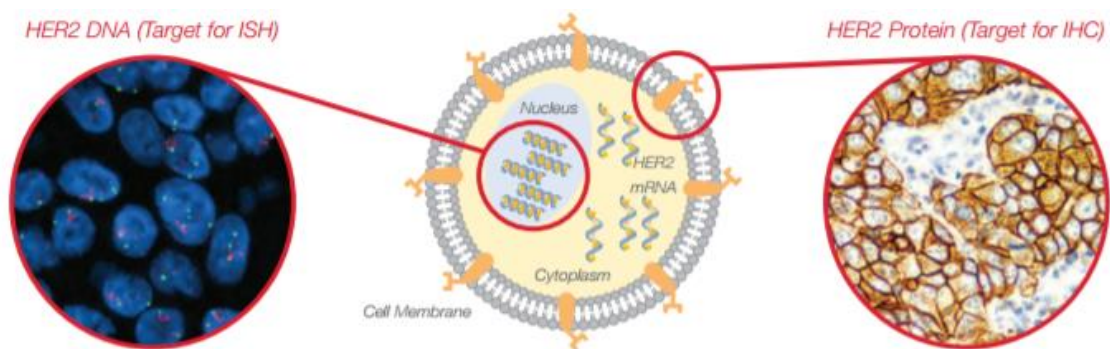


Figura 2.3: Pruebas de IHQ y FISH para detectar células cancerígenas por la presencia de HER2. En la imagen de la izquierda se observa pruebas a nivel genético de IF y a la derecha pruebas IHQ, revelando la presencia de la proteína, se observa membranas de la célula teñidas de color café. Fuente: www.agilent.com/cs/library/usermanuals/public/28630_herceptest_interpretation_manual-breast_ihc_row.pdf. Accedido el: 21-11-2016.

La ASCO/CAP, a través de sus metodologías ha propuesto un algoritmo de evaluación de HER2 mediante técnicas de IHQ, utilizando un proceso sistemático basado en la intensidad y completitud de membrana marcada [11, 19]. Los diversos pasos y etapas de dicho algoritmo se presentan en la Figura 2.4. Estas mismas organizaciones recomiendan el uso de algoritmos de análisis de imagen para la cuantificación del HER2 como un método deseable para la cuantificación objetiva de este marcador [6]. Sin embargo, a la fecha no se han desarrollado normas que especifiquen el rendimiento de estos sistemas automatizados, existiendo

simplemente algunas recomendaciones y lineamientos de referencia para introducirlos [41, 42].

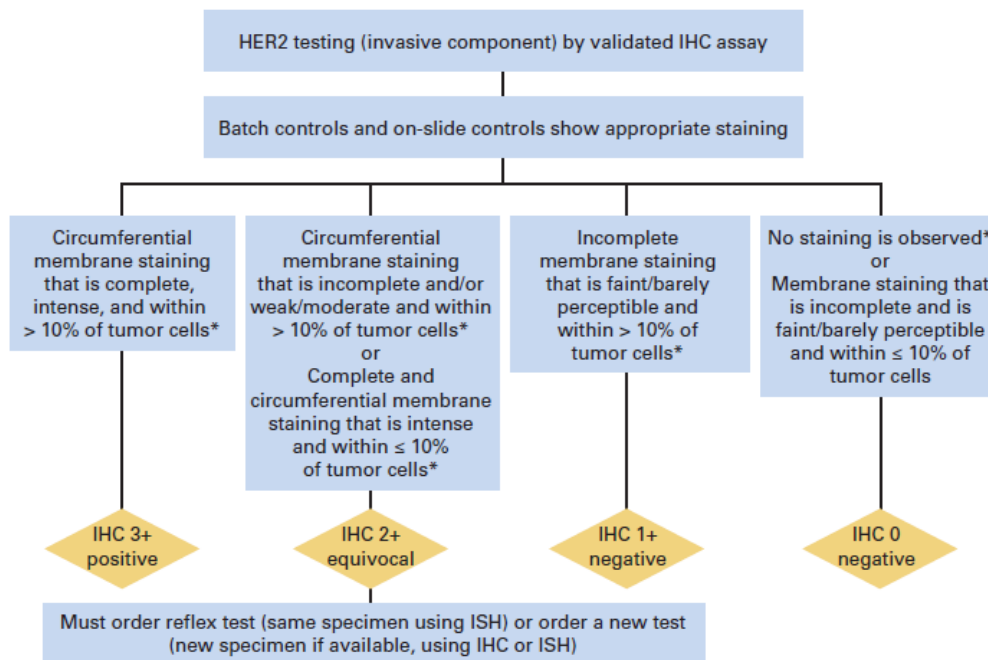


Figura 2.4: Algoritmo para la evaluación receptor 2 del factor de crecimiento epidérmico humano (HER2). Estados de expresión de la proteína presente mediante técnica IHQ de una biopsia de tejido mamario. Para casos equívocos (2+) se recomienda un análisis de hibridación inmunofluorescente in situ FISH. El resultado final de la evaluación asume que no hay discordancia histopatológica aparente observada e informada por el patólogo según su criterio de interpretación. Fuente: ascopubs.org/doi/full/10.1200/jco.2013.50.9984. Accedido el: 21-11-2016.

La sobreexpresión de la proteína HER2 en biopsias de tejido mamario es determinada tradicionalmente mediante estimación visual a través de un microscopio convencional, en la cual el patólogo entrega una clasificación, según los protocolos mencionados anteriormente, de la cantidad de inmunotinción (color café) presente en las células [19, 43] (Figura 2.5).

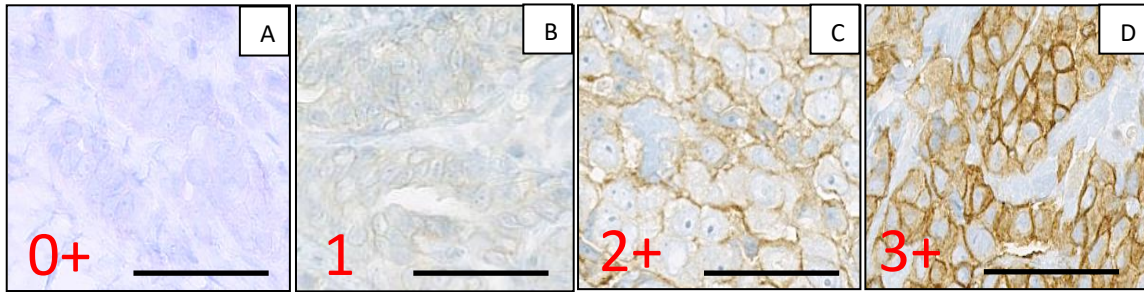


Figura 2.5: Ejemplo de la evaluación IHQ de HER2 según clasificación ASCO/CAP 0, 1+, 2+ y 3+. A) No se observa tinción de membrana celular en las células invasivas; B) Tinción incompleta de membrana celular en las células invasivas; C) tinción completa de la membrana celular pero no uniforme o tenue en intensidad en al menos el 10% de las células invasivas; D) Tinción completa de la membrana celular, uniforme e intensa en más del 10% de las células invasivas. Barra: 0,1 mm.

2.2.3 Consideraciones acerca de la evaluación de RE y RP

La importancia de los receptores de hormonas, estrógeno (RE) y progesterona (RP) en la biología del CM fue reconocida hace ya más de cuatro décadas, sobre la base de datos clínicos en humano, así como en animales¹. La detección de estos receptores se realiza mediante técnicas de IHQ que permiten reconocer su presencia o ausencia en el núcleo de la célula. La unión de los estrógenos a los RE estimula la proliferación de células mamarias, aumentando tanto el número de células diana dentro del tejido, así como la división celular y síntesis de ADN, lo que eleva el riesgo de ocurrencia de errores de replicación [12]. Por otro lado, existen estudios que sugieren que el gen que codifica para RP es estrógeno-dependiente, por lo que su sobreexpresión está estrechamente ligada al estrógeno. En este sentido, reportes dan evidencia que pacientes RE positivos y al mismo tiempo RP negativos presentarían un peor pronóstico comparado con pacientes RP positivos [12].

Desde el punto de vista científico, en la mayoría de los CM se expresan estos receptores y las células pueden ser estimuladas a crecer por las hormonas estrógeno y progesterona [40]. Por tanto, la eliminación de estas hormonas o el bloqueo de sus

¹ <https://www.uptodate.com/contents/hormone-receptors-in-breast-cancer-clinical-utility-and-guideline-recommendations-to-improve-test-accuracy>. Accedido el: 08-01-2016.

receptores podrían retrasar o prevenir el crecimiento tumoral y prolongar la supervivencia de la paciente. Aproximadamente entre el 75% a 80% de las mujeres con CM presentan RE y RP positivos [12, 40]. Clínicamente, el estado de estos receptores permite asignar a la paciente una terapia hormonal o endocrina, también conocida como terapia adyuvante [12, 40]. Este tratamiento ha probado ser eficaz y apropiado para casi todas las mujeres con RE y RP positivos, por lo que es ampliamente indicado para pacientes con CM, tanto en países desarrollados como en aquellos en desarrollo [44, 45].

2.2.4 Determinación de la positividad de RE y RP

Las guías y recomendaciones de la ASCO/CAP establecen por una parte considerar positivos los resultados de la evaluación de RE y RP si existe al menos un 1% de células tumorales con núcleos inmunoreactivos; por otra parte, se consideran negativos si existe menos del 1% de núcleos inmunoreactivos en células tumorales, basados en la evaluación de las características morfológicas de los núcleos y en la evidencia de invasión [12, 40]. El porcentaje puede ser determinado mediante i) estimación visual y/o cuantificación manual, realizado por médicos patólogos, o ii) mediante algoritmos de análisis de imagen (software). A continuación, se proporcionan algunos detalles técnicos de este último método [12, 40].

Los sistemas de cuantificación automatizada de RE y RP pueden usar la proporción o número de células positivas, o incluir la intensidad de la inmunoreactividad:

- Número de células positivas: El número de células positivas observadas en una biopsia puede ser reportado como un porcentaje dentro de una categoría discreta (Figura 2.6) [40].

Intensidad: Se refiere al grado de positividad nuclear, es decir la gradiente entre pálido a oscuro (Figura 2.7). Sin embargo, existen algunos factores que pueden afectar la intensidad; entre ellos se cuentan: i) cantidad de proteína presente, ii) cantidad de anticuerpo utilizado, iii) marca comercial o proveedor y iv) el sistema de recuperación del antígeno utilizado. En la mayoría de los CM, existe inmunoreactividad heterogénea de pálido a oscuro en las células positivas presentes [40].

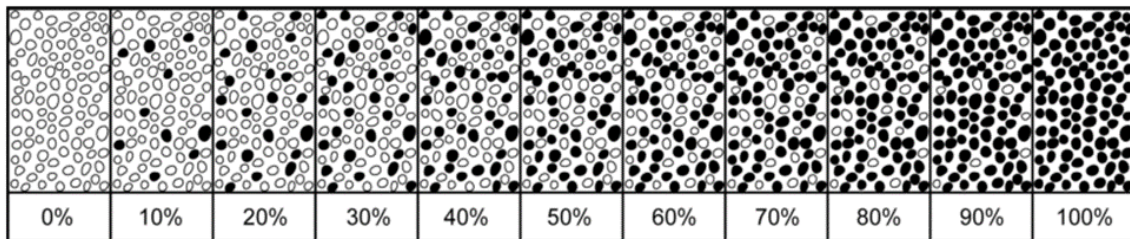


Figura 2.6: determinación del porcentaje de positividad nuclear en receptores hormonales. Ejemplo esquemático de la estimación visual del porcentaje de positividad. Fuente: www.cap.org/ShowProperty?nodePath=/UCMCon/Contribution%20Folders/WebContent/pdf/cp-breast-biomarker-template-14.pdf. Accedido el: 21-11-2016.

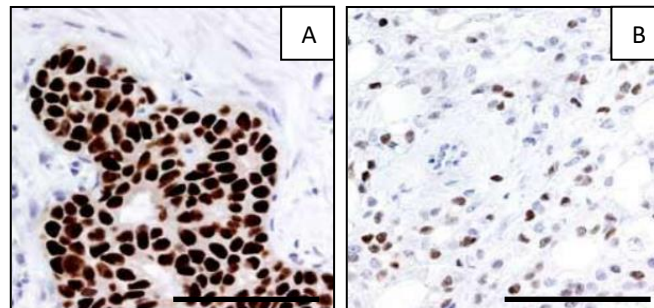


Figura 2.7: Ejemplo de la evaluación IHQ de RE según recomendación ASCO/CAP. A) Tinción nuclear mediante IHQ para RE con 100% de positividad en carcinoma ductal, se observa la marca IHQ en núcleos de color café oscuro; B) Ídem a A, RE 30% de positividad en carcinoma lobular, se observa núcleos marcados entre pálidos a oscuros. Barra: 2 mm. Fuente: www.nordiqc.org/downloads/assessments/13_2.pdf. Accedido el: 21-11-2016.

2.3 VARIABILIDAD EN LA EVALUACIÓN DE HER2 Y RECEPTORES DE HORMONAS

Sabiendo que el establecimiento de un diagnóstico preciso y definitivo permite decidir el tratamiento terapéutico más adecuado para el paciente, llama la atención que múltiples estudios han puesto en evidencia que la evaluación e interpretación de estos marcadores presenta una alta subjetividad. Más aún, tales estudios muestran que existe una considerable variabilidad inter e intra observador, lo que sin duda afecta directamente el resultado del diagnóstico. Entre los factores más importantes a los que se atribuye esta variabilidad se encuentran los procesos de IHQ [46-50].

La literatura indica que alrededor del 20% de las evaluaciones realizadas tanto para HER2 como para los receptores hormonales podrían ser imprecisas [11, 12, 15, 20]. Debido a la imprecisión de los resultados obtenidos en análisis de estos marcadores, es posible que mujeres diagnosticadas con CM no se beneficien con el tratamiento adecuado o queden expuestas a los riesgos de fármacos por tratamiento no adecuado [11, 12].

2.3.1 Control de calidad en Inmunohistoquímica - NordiQC

La IHQ es considerada una técnica especial y compleja, y como tal, sujeta a diversos errores en sus etapas pre-analíticas, analíticas y post-analíticas, tanto en su reproducibilidad como fiabilidad, los que pueden llegar a incidir en incertezas y diagnósticos equivocados [46, 50]. Este problema se encuentra directamente relacionado con el control de calidad de las muestras, debiéndose principalmente a la falta de estandarización universal y sistemática de protocolos de IHQ. Lo anterior se ve agravado debido a la producción y utilización de un número cada vez mayor de nuevos anticuerpos y reactivos [16, 50].

En varios países han surgido instituciones que promueven la realización de controles de aseguramiento de la calidad externa, ofreciendo evaluaciones de competencias a los laboratorios que desarrollan procesos de IHQ. En dichas evaluaciones, la calidad de las muestras es evaluada en base a una serie de criterios

protocolizados y estandarizados sistemáticamente, ofreciendo, en caso de existir falencias, sugerencias específicas tendientes a mejorar protocolos y procedimientos [50, 51].

Una de estas instituciones es “*Nordic immunohistochemical Quality Control*” (NordiQC), una organización científica profesional con casa matriz en el Instituto de Patología del Hospital Universitario de Aalborg (Dinamarca)¹. NordiQC surgió en respuesta a la necesidad de homogeneizar los procesos de IHQ, debido a la existencia de variaciones considerables en las normas y rutinas seguidas en diversos laboratorios [50, 52]. El 2003 NordiQC llevó a cabo la primera evaluación externa de calidad donde participaron 71 laboratorios, los resultados evidenciaron que alrededor del 50% de los laboratorios participantes presentaban tinciones insuficientes (Figura 2.8A) [53]. En la actualidad esta institución gestiona y realiza evaluaciones externas de control de calidad de forma independiente a laboratorios procedentes de más de 60 países [50]. Su objetivo es entregar una evaluación del estado de variabilidad inter-laboratorios, propiciando las buenas prácticas, mejorando la comunicación y creando una instancia de consulta entre ellos [50].

¹ <http://www.nordiqc.org/about.php>. Accedido el: 08-01-2016.

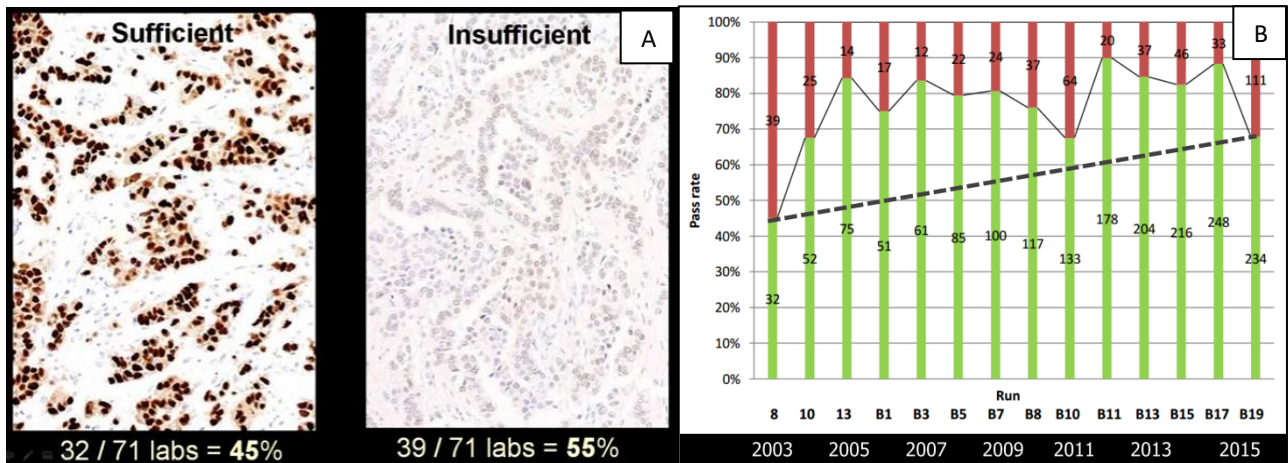


Figura 2.8: Resultados del control de calidad externo realizado por NordiQC. A) Comparación del tratamiento IHQ de un mismo tejido para RE por dos laboratorios (A y B). Se observa la falta de concordancia entre ambos resultados. Los resultados de la evaluación realizada el 2003 con la participación de 71 laboratorios europeos, evidenciaron que solo el 45% de los laboratorios participantes presentaban tinciones suficientes, el restante 55% tinciones insuficientes debido a factores pre-analíticos, analíticos como tiempos de secado, marcas de anticuerpos, entre otros ocasionando diagnósticos equívocos; B) Grafico de barras del número de laboratorios participantes entre los años 2003 a 20015, en verde destacan los laboratorios que tuvieron tinciones suficientes, en rojo, laboratorios que presentaron tinciones insuficientes. Se observa el rendimiento de los laboratorios influenciada por la armonización del protocolo sugerida por NordiQC. Fuente: <https://www.yumpu.com/en/document/view/47592457/small-nordiqc>. Accedido el: 21-11-2016.

En Chile, un estudio realizado el año 2012 sobre la variabilidad del análisis del marcador de HER2 [20], concluyó que, “la variabilidad de la evaluación del estado de HER2 en laboratorios nacionales es similar a lo descrito en la literatura internacional, llegando al 19,7% cuando se compara con la técnica de IHQ estandarizada y alcanza el 26,9% cuando se compara con la técnica de inmunofluorescencia FISH”. Sin embargo, este estudio no identifica las fuentes de variabilidad de HER2 [20]. En la actualidad, considerando las nuevas herramientas tecnológicas disponibles, como son la digitalización de placas, el acceso a placas virtuales mediante internet y el análisis de imágenes, es plausible proponer que estas herramientas de PD podrían ayudar a detectar las fuentes de variabilidad o los factores que influyen en la evaluación de manera mucho más específica y precisa.

2.4 PATOLOGÍA DIGITAL

La Patología Digital (PD) permite la adquisición de imágenes gestión e interpretación de la información proveniente desde la solicitud de análisis de una biopsia hasta la entrega del informe histopatológico, proceso que incluye el análisis de un portaobjeto de vidrio digitalizado denominado placa virtual [1, 3, 6, 54].

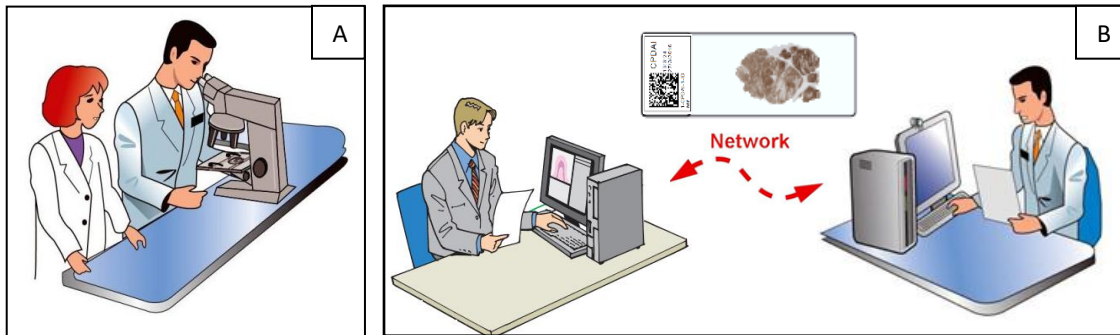


Figura 2.9: Transformación de un flujo de trabajo manual a un flujo de trabajo digital. A) Se observa el análisis y evaluación de biopsias según el método tradicional. B) Entorno dinámico basado en imágenes (placas virtuales) y conectividad a través de Internet. Fuente: www.hamamatsu.com/resources/pdf/sys/SBIS0065E_NDPSEERVE.pdf. Accedido el: 21-11-2016.

La PD proporciona un entorno para presentar, transmitir, archivar y transportar información (datos e imágenes) de patología mediante plataformas informáticas y tecnologías de información (Figura 2.9). Su utilización reduce la necesidad de desplazar físicamente portaobjetos de vidrio de un patólogo a otro o desde el laboratorio clínico a un patólogo, y tiene por finalidad apoyar el flujo de trabajo desde la recepción de muestras hasta la emisión de un informe [1, 6, 54].

2.4.1 Placa Virtual y su almacenamiento

La placa virtual es un archivo de datos de gran tamaño resultante de la digitalización de un portaobjeto o placa de vidrio histopatológico digitalizado a través de un escáner de tejido [1]. Estos escáneres son microscopios controlados por software que capturan un conjunto de imágenes secuenciales, las que son posteriormente unidas en forma automática creando una placa virtual [1, 3, 6, 54]. La placa virtual es un archivo de imagen que representa una réplica del portaobjeto digitalizado en alta resolución. Este archivo puede ser visualizado de forma remota

mediante el uso de diversos softwares de visualización, que son plataformas informáticas que emulan las funcionalidades de un microscopio convencional [1], permitiendo navegar por el tejido a distintos aumentos, realizar anotaciones, mediciones, lo que presenta numerosas ventajas en las áreas académica, clínica e investigación [1-3].

Los archivos de imagen son del orden de 10 billones de píxeles, para una magnificación de 40x, sin *z-stacks*, por lo que requieren un espacio en memoria RAM del orden de los 30 a 56 gigabytes por placa virtual en formato sin comprimir. Es evidente que a esta escala el manejo de datos no es posible de realizar con la capacidad de un solo computador [55, 56]. Por lo tanto, su manipulación requiere de infraestructura tanto en *software* y *hardware* que permita manejar este tipo de imágenes. Con el fin de abordar estos desafíos de manejo de datos, actualmente la computación de alto rendimiento se ha empleado tanto para almacenamiento como para aplicaciones de análisis de imágenes de patología, en ámbitos comerciales y de investigación [57].

2.4.2 Análisis de imagen

El análisis de imagen (AI), es un proceso que implica la segmentación de objetos, regiones, extracción de características, selección y etapas de clasificación, aplicados particularmente en el diagnóstico asistido por computador [2, 6]. Hoy en día es uno de los objetivos principales de la PD para generar datos cuantitativos y/o cualitativos a partir de placas virtuales. La idea básica del AI es utilizar una serie de algoritmos matemáticos para detección de patrones histopatológicos, basados en color, forma, textura y/o contexto [6]. Actualmente el uso de placas virtuales ocasionó un incremento en investigadores y empresas desarrolladoras de *software* para la creación de algoritmos de cuantificación de patrones histopatológicos, con énfasis en el cáncer.

Es así como en CM existen diversos algoritmos para detección y cuantificación de marcadores tumorales, entre los que destacan aplicaciones para HER2, RE y RP debido a la relevancia de su análisis [5, 7]. En el mercado existen *softwares*

comerciales para análisis de imagen en placas virtuales, que ofrecen diversos algoritmos orientados hacia distintos tipos de cáncer^{1,2}. Entre ellos se encuentra la empresa danesa Visiopharm, cuyo *software* ha sido exitosamente instalado en 7 hospitales europeos (Hospital Universitario de Herlev, Aarhus, Aalborg, etc.), para su uso clínico y de investigación^{3,4}.

2.4.3 Software comercial de análisis de imágenes – Visiopharm

La empresa Visiopharm, caracterizada por una propuesta innovadora de algoritmos de aplicación en cáncer, ha creado diversos paquetes de software dedicados al análisis de imagen. Entre ellos se encuentra el NDP.analyze cuya plataforma integra herramientas de análisis de imagen denominada VIS (*Visiopharm Integrator System*), permitiendo la adquisición de imágenes, control de hardware (cámara), base de datos e interface usuario. En dicha plataforma se arman y ejecutan módulos específicos, tales como: VisiomorphDP (herramientas de análisis cuantitativo, propiedades espectrales, espaciales, morfológicas, entre otras), Tissuemorph (detección nuclear) y HER2-CONNECT (detección de membrana celular). A través de estos módulos el usuario puede crear o configurar algoritmos de cuantificación para diferentes patrones biológicos. La plataforma cuenta con procesos de análisis (llamados también ‘Apps’) que son algoritmos básicos pre-definidos para detección de citoplasma, núcleos y membrana. Para la aplicación específica y adecuada de marcadores tumorales, estos algoritmos requieren de etapas de configuración y calibración local. Por lo tanto, se hace necesario realizar una meticulosa estandarización y validación local de estos algoritmos [58].

¹ <https://www.visiopharm.com/>. Accedido el: 16-11-2016

² <http://www.definiens.com/>. Accedido el: 16-11-2016

³ <http://www.prweb.com/releases/visiopharm/pressreleases/prweb11855468.htm>. Accedido el 15-06-2015

⁴ <https://www.visiopharm.com/>. Accedido el 15-02-2017

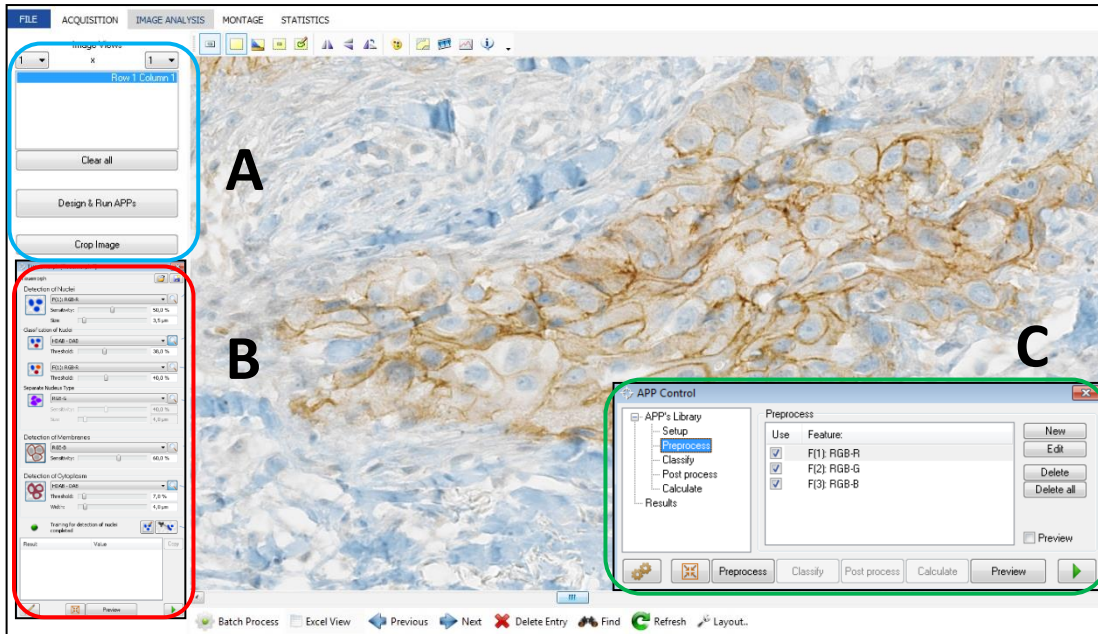


Figura 2.10: Plataforma de trabajo del software NDP.analyze de análisis de imagen de Visiopharm. A) Menú de las herramientas principales e indicador de módulos instalados; B) Ventana de las herramientas del módulo Visiomorph para control de algoritmos de membrana, nuclear y citoplasma. Permite la selección manual de parámetros de sensibilidad; C) Ventana de configuración de algoritmos, herramientas de pre-procesamiento, clasificación y post-procesamiento.

2.4.4 Algoritmo para detección de HER2 en membrana celular

El algoritmo de Visiopharm específico para detección de HER2 en membrana celular se encuentra desarrollado dentro del módulo HER2-CONNECT e incluye las etapas de:

- i) pre-procesamiento, para detección de píxeles de color café dentro de estructuras lineales (membrana celular) determinados por la presencia de HER2 (característica particular) en secciones de tejido procesados en IHQ;
- ii) segmentación bimodal, para distinguir píxeles que representan la membrana específicamente teñida, de otros píxeles inespecíficos dentro de la ROI;
- iii) post-procesamiento, para eskeletonizar la membrana, discriminando membranas no perfectamente conectadas y eliminando fragmentos muy pequeños de membrana [4, 59].

Los valores de los parámetros y variables utilizados en el pre-procesamiento, segmentación y post-procesamiento se establecieron en un estudio precedente con la institución NordiQC (Hospital de Aalborg, Dinamarca) utilizando diferentes métodos de tinción y otro modelo de escáner de placas histológicas (Nanozoomer HT 1.0), entre otros [4, 59]. Sin embargo, para propósitos de estandarización y configuración local, el algoritmo requiere de 2 parámetros de entrada: el primero es la ROI entregada por el patólogo y el segundo es la sensibilidad definida por el usuario [4, 58]. Este último parámetro tiene un rango entre 0% y 100% y el usuario debe definir uno o más valores que le permitan obtener una mejor detección de la membrana marcada (completitud de la membrana). En este contexto cabe hacer notar que el algoritmo reconoce automáticamente la membrana marcada, no requiere de un proceso de entrenamiento adicional. Como resultado, entrega un parámetro denominado conectividad, que es una medida cuantitativa de la estructura de la membrana marcada. La conectividad oscila entre 0 y 1, donde cero indica que no hay membrana marcada y 1 que la membrana está completamente marcada, presentando un aspecto característico denominado *Chicken wire* [4, 59] (Figura 2.11). Este valor numérico de la conectividad de membrana marcada entregado por el algoritmo requiere ser traducido a la nomenclatura tradicional (0/1+, 2+ y 3+), en este contexto los desarrolladores utilizaron el clasificador de CART (*Classification and Regression Trees*) que viene acoplado al módulo HER2-Connect [4], sin embargo, la licencia de uso adquirida por CPDAI (licencia del tipo investigación), no viene con el clasificador incorporado. Por tanto, para este propósito se requiere la incorporación de un método de traducción.

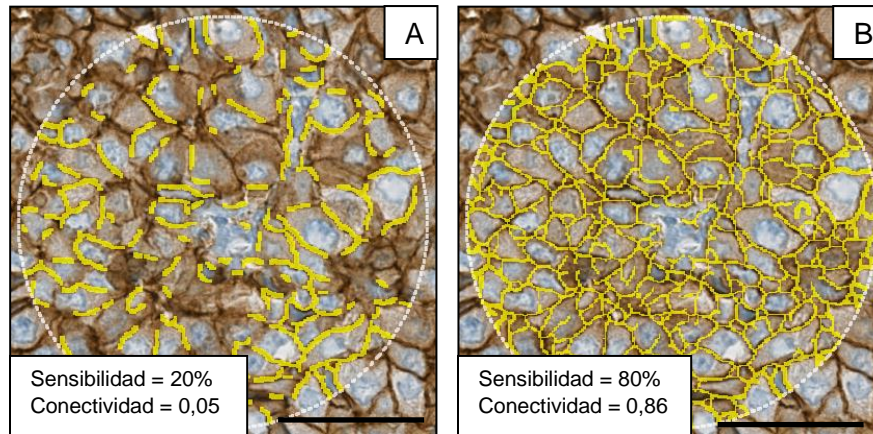


Figura 2.11: ROI de una biopsia mamaria analizada por el algoritmo HER2 configurado bajo dos valores distintos de sensibilidad: El círculo blanco punteado denota la ROI seleccionada por el patólogo, sobre la cual se aplicó posteriormente el algoritmo configurado a un 20% (A) y 80% (B) de sensibilidad, arrojando sus respectivos valores de conectividad. Las líneas amarillas indican la membrana HER2 positiva detectada por el algoritmo. A) Con sensibilidad a 20%, entrega un índice de conectividad de 0,05; B) Con sensibilidad a 80% el índice de conectividad alcanza un 0.86. Barra: 0,25 mm.

Este algoritmo no solo se basa en la intensidad de tinción, sino que además considera la continuidad de membrana celular marcada, lo que permite obtener un mayor rendimiento en términos de detección y cuantificación de la membrana por la presencia de HER2, esta metodología de detección de membrana obtuvo mayor rendimiento y precisión comparado a algoritmos que sólo se basan en la intensidad de tinción [4, 7].

2.4.5 Algoritmo para detección nuclear de RE y RP

Para la detección de núcleos el algoritmo de Visiopharm identifica características específicas tales como color, textura y tamaño a través de un proceso de entrenamiento previo, con el fin de definir previamente tales características [58]. Para este propósito se utiliza un método de aprendizaje en el cual se entregan instrucciones al algoritmo para que identifique o detecte particularidades de los núcleos. En el caso de receptores hormonales se especifica la diferenciación de los núcleos positivos respecto a los negativos, discriminando el tejido de fondo [58] (Figura 2.12).

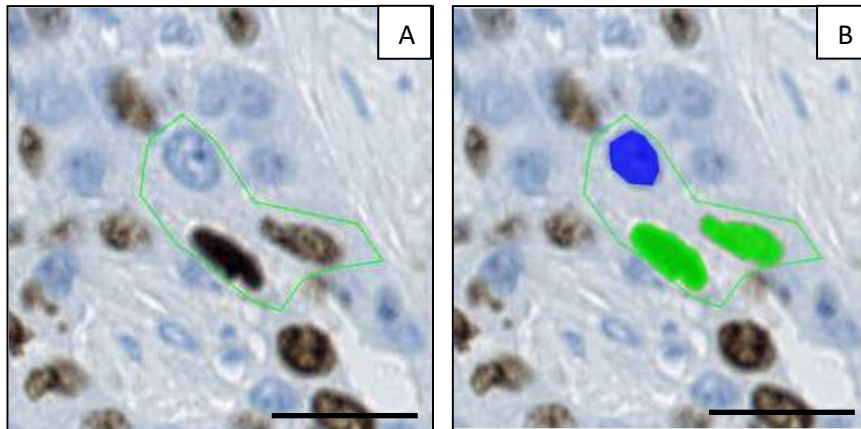


Figura 2.12: Selección de núcleos de entrenamiento en una muestra para RE. A) Delimitación de la región de interés a analizar; B) Selección manual de núcleos positivos (verde) y negativos (azul) para el entrenamiento del algoritmo. Barra 0,2 mm. Fuente: User Manual NDP.analyze (Visiopharm|Hamamatsu).

El algoritmo consta de etapas de pre-procesamiento, segmentación y post-procesamiento que se describen a continuación.

En la etapa de pre-procesamiento, el algoritmo usa una banda de deconvolución de color (HDAB-DAB) para detectar núcleos teñidos positivamente (color café, característica particular) y una multiplicación de la banda de color rojo y azul para detectar núcleos negativos (color azul) [58]. Para la segmentación utiliza un método de separación de núcleos basado en la probabilidad de forma y tamaño, empleando una técnica de segmentación de núcleos denominada *watershed* [60], que también puede ser aplicada para la clasificación de núcleos positivos según niveles de intensidad (débil, moderado y fuerte). En el post-procesamiento, las áreas de núcleos demasiado pequeñas ($< 5 \mu\text{m}$; pero pueden ser redefinidas por el usuario) son eliminadas. La imagen obtenida después del post-procesamiento se utiliza para determinar las variables de salida, es decir, número de núcleos, tanto positivos como negativos, así como los porcentajes de positividad, entre otros [58].

2.5 CLASIFICADORES DE APRENDIZAJE AUTOMÁTICO

El aprendizaje automático (*Machine Learning*, ML) es una rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que, mediante el diseño y desarrollo de

algoritmos permitan a las computadoras aprender. La información requerida para esto es suministrada en conjuntos de ejemplos (muestras de datos) de modo de alcanzar un concepto general de inferencia [61]. Actualmente existe un creciente número de sistemas y dominios de aplicación en medicina, especialmente en áreas donde el conocimiento es difícil de adquirir manualmente. Los enfoques actuales hacen uso de datos reales y modelos de dominio, es decir, conjunto de valores o atributos, para impulsar el proceso de adquisición del conocimiento [61]. ML también puede ser utilizada para perfeccionar bases de conocimientos existentes entre el conocimiento del dominio proporcionado por el experto y los casos reales [61]. Los diferentes algoritmos de ML se agrupan dentro de una taxonomía dependiendo de la función de salida de los mismos. A continuación, se explican brevemente las técnicas de los algoritmos de ML utilizados en la presente tesis para llevar a cabo la clasificación automática del estado de HER2 en CM.

2.5.1 Árboles de decisión - CART

Los árboles de decisión (*Decision Trees*, DTs) son algoritmos de ML comúnmente usados para clasificación y predicción, cuyo atractivo se debe al hecho que presentan reglas específicamente estructuradas. Estas reglas - o condiciones - son fácilmente expresadas, de manera que los seres humanos pueden entenderlas o incluso utilizarlas directamente en un lenguaje de acceso a bases de datos [62]. Un DTs tiene una estructura similar a un diagrama de flujo, donde cada nodo denota una prueba en un atributo, cada rama representa el resultado de una prueba y cada hoja (o terminal) tiene asociada una etiqueta de clase que corresponde al rango de clasificación final. El nodo superior en un árbol es el nodo raíz (Figura 2.13).

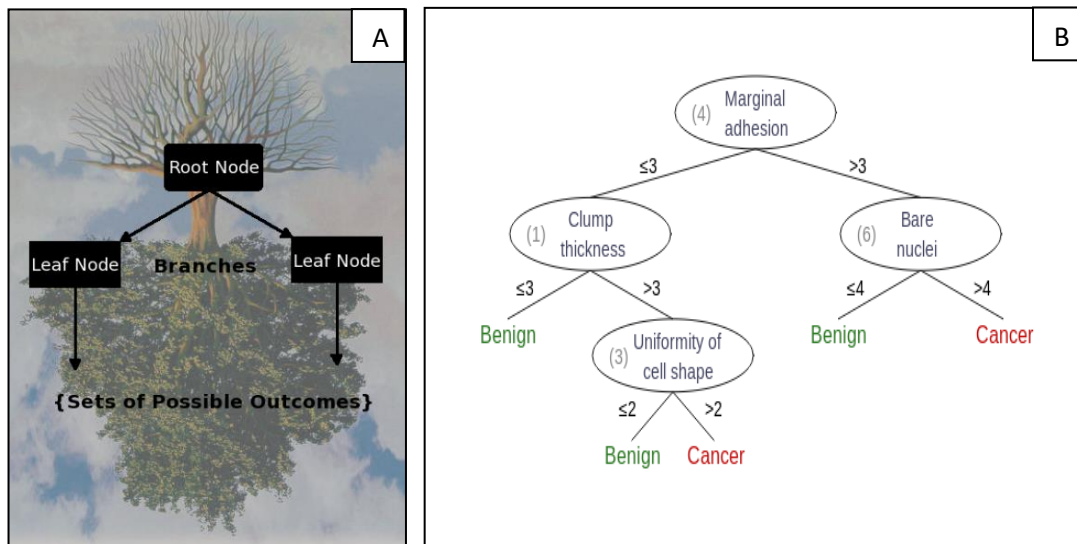


Figura 2.13: Algoritmo de árboles de decisión en máquinas de aprendizaje. A) Componentes de un árbol de decisión; B) Esquema de un árbol de decisión. Fuente: www.artofcomputing.cecs.anu.edu.au/images/lab8/dt-sample.png. Accedido el: 21-11-2016.

Los modelos de un DTs donde la variable objetivo toma un conjunto finito de valores (variables categóricas) son llamados árboles de clasificación, en este caso las hojas representan las etiquetas de clase y las ramas representan la conjunción de características que conducen a esas etiquetas de clase. En los casos donde la variable objetivo puede tomar valores continuos (típicamente números reales) se llaman árboles de regresión¹.

El análisis CART (*Classification and Regression Trees*) utiliza un particionamiento recursivo para crear DTs, donde cada nodo representa una celda de partición [63], construyendo árboles de clasificación y regresión para predecir variables dependientes continuas (regresión) y variables predictivas categóricas (clasificación) [64].

¹ https://en.wikipedia.org/wiki/Decision_tree_learning. Accedido el: 20-12-2016

2.5.2 Máquinas de Vectores de Soporte – SMO

Las máquinas de vectores de soporte (*Support Vector Machines*, SVM) son otro tipo de técnica de ML¹, mediante la cual se construye un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad muy alta, con el propósito de maximizar el margen o separación entre dos clases [65]. El SVM es considerado como un clasificador muy preciso y de aplicación en muchas áreas y puede ser utilizado en problemas de clasificación o regresión [66].

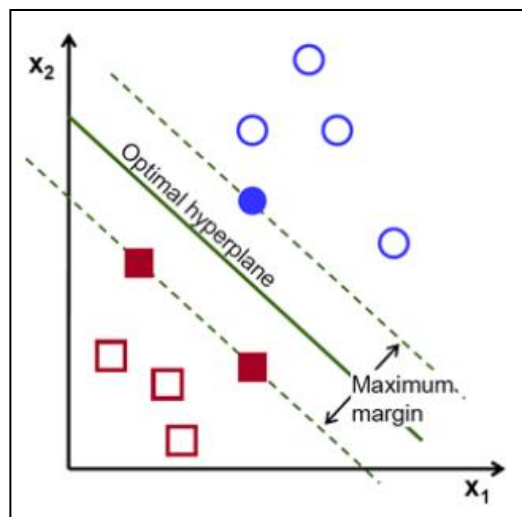


Figura 2.14: Ilustración simplificada de una clasificación lineal mediante SVM. Datos de entrada clasificados y separados a través de un hiperplano. Fuente: www.quora.com/What-does-support-vector-machine-SVM-mean-in-laymans-terms. Accedido el: 21-11-2016.

Generalmente, el entrenamiento tradicional del SVM presenta un problema de optimización cuadrática (minimiza o maximiza una función cuadrática de varias variables de decisión sujetas a restricciones lineales)² [67]. El SMO es un algoritmo de SVM que soluciona este problema mediante su división en sub-problemas más pequeños. Estos sub-problemas se resuelven analíticamente, evitando las iteraciones internas de los algoritmos de optimización, es decir, no requiere de una

¹ https://forecasters.org/wp-content/uploads/gravity_forms/7-621289a708af3e7af65a7cd487aee6eb/2015/07/ISF2015_PPT.pdf. Accedido el: 20-12-2016

² https://en.wikipedia.org/wiki/Quadratic_programming. Accedido el 20-12-2016

matriz extra de almacenamiento, ni invocar rutinas numéricas de optimización [67, 68]. En general el SVM es considerado como un clasificador muy preciso. En esta tesis se optó por su uso debido al tamaño relativamente acotado de la muestra, a que es fácil de implementar y al hecho que ha sido previamente utilizado en otros estudios relacionados con cáncer [61, 66, 67, 69].

2.6 MODELADOR DE PROCESOS

La gestión del conocimiento médico y la gestión de procesos son considerados un activo valioso en el ámbito de la salud, ya que la naturaleza de los procesos asistenciales depende en gran medida de una interacción multidisciplinaria para identificar, revisar, validar, representar y comunicar el conocimiento a personas, sistemas y perspectivas de la organización [70]. Entre la gran variedad de lenguajes de modelamiento de procesos, el Modelo y Notación de Procesos de Negocio (*Business Process Model and Notation, BPMN*) es considerado un estándar de facto según la norma ISO/IEC 19510:2013¹. Esta norma tiene por fin proporcionar una notación fácil y comprensible para todos los usuarios, desde analistas del negocio hasta los desarrolladores técnicos y responsables de la aplicación de la tecnología, y

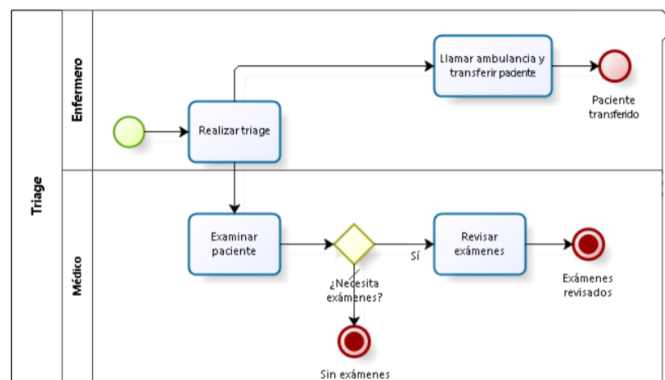


Figura 2.15: Modelado del proceso de triaje en un centro médico. Se ejemplifica el modelado del proceso, las actividades y acciones a realizar. El proceso inicia con la información básica de la salud del paciente. De acuerdo a los exámenes requeridos el usuario puede registrar la transferencia de un paciente a otro centro médico. Fuente: www.bizagi.com/processxchange/Documents/bfa6771d-a0b1-4eda-b342-748466490aa4/docs/Sala%20de%20Emergencias-Descripci%C3%B3n.pdf. Accedido el: 12-05-2015

¹ http://www.iso.org/iso/catalogue_detail.htm?csnumber=62652. Accedido el: 12-05-2015

finalmente a los usuarios del servicio o negocio [70].

El diseño es claro y fácil de seguir para usuarios no técnicos y lo convierte en un candidato prometedor para representar procesos en el área de la salud, donde el personal médico necesita entender y discutir los modelos de procesos asistenciales¹ (Figura 2.15) [70-72].

2.7 ESTADÍSTICOS DE CONCORDANCIA Y NORMALIDAD.

Las calificaciones o evaluaciones asignadas a los mismos sujetos o muestras pueden diferir mucho de un evaluador a otro. Esta variación entre evaluadores puede poner en peligro la integridad de las investigaciones científicas y/o tener un impacto sesgado en los resultados entregados. De hecho, un fármaco equivocado o una dosis incorrecta pueden llegar a ser administrados a pacientes en un hospital debido a un diagnóstico equivoco. Estos problemas fueron conducentes al estudio de confiabilidad entre evaluadores [72, 73].

2.7.1 Coeficiente Kappa de Cohen (k)

La variabilidad inter-observador, se evalúa a través de pruebas de concordancia, cuyo objetivo es estimar hasta qué punto dos observadores coinciden en su medición. Este problema es abordado según la naturaleza de los datos [73]. Cuando éstos son de tipo categórico o nominal como en el caso de las clasificaciones del estado de HER2 (0/1+, 2+ y 3+) el análisis más frecuente es la prueba Kappa (k) de Cohen, cuyo coeficiente k refleja la fuerza de la concordancia existente entre dos evaluadores, tomando valores entre -1 y +1. Valores cercanos a +1, mayor es el grado de concordancia [73, 74]. Por el contrario, mientras más cercano a -1, mayor es el grado de discordancia. 0 representa la concordancia observada que se espera a causa exclusiva del azar [74]. La fórmula se expresa de la siguiente forma:

¹ <http://www.bizagi.com/processxchange/Documents/bfa6771d-a0b1-4eda-b342-748466490aa4/docs/Sala%20de%20Emergencias-Descripci%C3%B3n.pdf>. Accedido el: 12-05-2015

Ecuación 1: Formula para el cálculo de concordancia Kappa de Cohen

$$\kappa = \frac{[(\Sigma \text{concordancias observadas}) - (\Sigma \text{concordancias atribuibles al azar})]}{(\text{total de observaciones}) - (\Sigma \text{concordancias atribuibles al azar})}$$

En la Tabla 2.1 se presenta la escala que expresa cualitativamente la fuerza de concordancia.

Tabla 2.1: Niveles de concordancia del coeficiente Kappa de Cohen. Valoración del nivel de concordancia propuesta por Cohen.

Kappa (k)	Nivel de concordancia
0,00	Pobre (<i>Poor</i>)
0.01 - 0.20	Leve (<i>Slight</i>)
0.21 - 0.40	Aceptable (<i>as fair</i>)
0.41 - 0.60	Moderada (<i>Moderate</i>)
0.61 - 0.80	Considerable (<i>Substantial</i>)
0.81 - 1.00	Casi perfecta (<i>Almost Perfect</i>)

2.7.2 Coeficiente Alpha de Krippendorff (α)

El coeficiente de Alpha Krippendorff (α) mide el acuerdo o concordancia entre observadores, codificadores, jueces, evaluadores o mediciones de instrumentos, realizando distinciones entre fenómenos no estructurados, asignando valores computacionales a ellos [75]. Se aplica α por las características de nuestros datos como es el caso de los porcentajes de positividad en receptores hormonales. Este coeficiente, generaliza varios índices de confiabilidad, permitiendo el juicio de varios datos con el mismo estándar de confiabilidad [77]. Puede ser usado para distintos tipos de números, que pueden ser categóricos, valores de escala, unidades de

medida (nominal, ordinal, intervalo, proporción, etc.) [76-78]. La fórmula de α es la siguiente:

Ecuación 2: Formula para el cálculo del coeficiente α de Krippendorff

$$\alpha = 1 - \frac{Do \text{ (desacuerdo observado)}}{De \text{ (desacuerdo esperado)}}$$

Cuando los observadores concuerdan se tiene que $Do = 0$ y $\alpha = 1$, lo que indica una confiabilidad perfecta. Mientras una concordancia por azar genera e $Do = De$ y $\alpha = 0$, lo que indica la ausencia de confiabilidad. En la Tabla 2.2 se presentan rangos para expresar Niveles de confiabilidad.

Tabla 2.2: Niveles de confiabilidad de Krippendorff alpha. Valoración de la fuerza de concordancia propuesta por Landis y Koch, 1997.

K-alpha	Nivel de confiabilidad
$\alpha < 0.667$	No confiable
$0.800 > \alpha \geq 0.667$	Considerable
$\alpha > 0.800$	Confiable

2.7.3 Prueba de normalidad Shapiro–Wilk

La prueba de Shapiro-Wilk mide la normalidad de un conjunto de datos¹. Utiliza el contraste de hipótesis para rechazar la normalidad de la muestra. La hipótesis nula asume que la muestra proviene de una población distribuida normalmente. Si el valor p-value es menor al nivel de significancia establecido ($p < 0.05$) se rechaza la hipótesis nula y se considera que hay evidencia para concluir que la muestra no proviene de una distribución normal [78]. La normalidad del conjunto de datos puede

¹ https://es.wikipedia.org/wiki/Test_de_Shapiro%E2%80%93Wilk. Accedido el: 12-01-2016

ser contrastado visualmente, mediante gráficos de cuantiles normales (Q-Q), gráficos de cajas e histogramas [79].

2.7.4 Prueba de Tukey

La prueba de Tukey - HSD (*Honest Significant Difference*, HSD) es una prueba de comparaciones múltiples¹. Permite testear si los datos de un grupo se encuentran igualmente dispersos respecto a otro grupo. Esta prueba trata de especificar la hipótesis alternativa genérica como la de cualquiera de las pruebas de ANOVA (pruebas paramétricas). Se basa en la distribución que sigue la diferencia del máximo y del mínimo de las diferencias entre la media muestral y la media poblacional de un número de variables normales independientes e idénticamente distribuidas [80].

2.7.5 Prueba de Kolmogorov-Smirnov

La prueba de Kolmogórov-Smirnov (K-S) es una prueba no paramétrica que determina la calidad de ajuste de dos distribuciones de probabilidad entre sí, válido para variables aleatorias continuas. Las pruebas tienen por objetivo determinar si los datos de una población se ajustan a distribución normal², es decir contrastan la hipótesis de normalidad de una población.

¹ https://en.wikipedia.org/wiki/Tukey's_range_test. Accedido el: 08-03-2016

² <http://www.uv.es/~friasnav/SupuestosParametrica.pdf>. Accedido el: 08-03-2016

3.MATERIALES Y METODOS

El trabajo realizado en esta tesis formó parte del hito “Estandarización mediante investigación y desarrollo (I+D) de rutinas matemáticas para la cuantificación IHQ de membrana celular y nuclear en biopsias de tejido mamario humano”, del proyecto FONDEF D11I1096¹. Se utilizó un escáner de placas histológicas Nanozoomer XR (Hamamatsu, Japón), el software comercial de análisis de imagen NDP.analyze (Visiopharm, Dinamarca) y visualizadores *on-line* de placas virtuales, aplicadas a CM.

La metodología utilizada constó de cuatro etapas, las que se ilustran en la Figura 3.1. Cada una agrupó actividades y requerimientos necesarios para el cumplimiento de uno o más de los objetivos planteados, los que han sido descritos en el Capítulo 1, Sección 1.4.

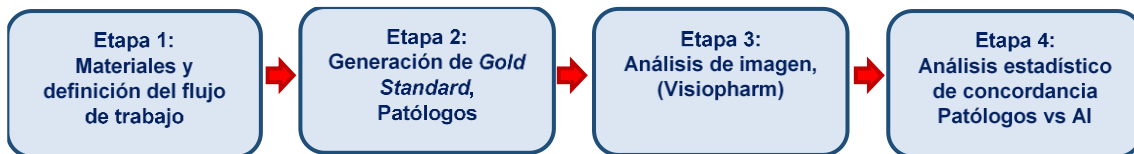


Figura 3.1: Etapas del proceso de validación de PD en CM. Etapa 1) Los materiales y recursos utilizados determinan las actividades y pasos a seguir durante el resto del proceso. Etapa 2) Evaluaciones realizadas por los patólogos constituyen el *gold standard* que se utilizará posteriormente en la cuantificación automatizada. Etapa 3) Da cuenta de dos objetivos: la configuración de los algoritmos y su aplicación en la cuantificación de los mismos casos que fueron evaluados por los patólogos en la etapa 2. Etapa 4) Comparación de los resultados de la evaluación realizada por los patólogos versus la cuantificación obtenida mediante los algoritmos.

3.1 ETAPA 1: MATERIALES Y DEFINICIÓN DEL FLUJO DE TRABAJO

Esta etapa establece para la planificación del proceso, así como su posterior diseño e implementación. En base a la literatura investigada, se definieron requisitos y se gestionaron los materiales, adaptando métodos y elementos a la realidad local con el propósito de diseñar e implementar el proceso de validación de PD en CM, con énfasis en la estandarización y validación de algoritmos de cuantificación utilizados para los marcadores tumorales HER2, RE y RP. A continuación, se

¹ <http://www.microscopiavirtual.cl/wp/>. Accedido el: 08-12-2016

presentan los requisitos necesarios para definir el flujo de trabajo y su posterior implementación.

3.1.1 Biopsias procesadas con IHQ y adquisición de placas virtuales

Para nuestro estudio se utilizaron 46 placas histológicas de biopsias core de CM invasivo procesadas mediante IHQ. Las placas fueron provistas por el Servicio de Anatomía Patológica del Hospital Clínico de la Universidad de Chile. El trabajo de tesis fue evaluado y aprobado por el Comité de Ética de Investigación en Seres Humanos (CEISH) de la Facultad de Medicina de la Universidad de Chile (ver Acta de autorización Anexo I).

Las placas provistas bajo una numeración interna del Servicio fueron registradas en CPDAI a través de un código sistemático con el fin de conocer solo su diagnóstico asociado, de modo que no se tuvo acceso a ficha clínica de paciente, ni datos del mismo. De estas placas, 26 casos correspondieron a HER2, 10 a RE y 10 a RP. Dichos casos fueron seleccionados por una patóloga experta del Servicio. El listado completo de los casos seleccionados y sus diagnósticos se presentan en el Anexo II. Todo el proceso de ingreso y escaneo de las placas se llevó a cabo bajo un riguroso sistema de trazabilidad, desarrollado en CPDAI con el fin de identificar y dar seguimiento específico a cada placa y evitar posibles pérdidas¹.

Las muestras fueron digitalizadas con el *Digital Slide Scanner* Nanozoomer XR, configurado en modo campo claro a 40x. El escáner genera archivos de imagen con formato ndpi denominados placas virtuales, los que fueron almacenados en servidores con alta seguridad y capacidad enlazados a la red CPDAI² – BioMed–HPC3. Las carpetas y placas virtuales fueron distribuidas a los patólogos a través de internet mediante la plataforma de visualización NDP.serve–NDP.view (Hamamatsu, Japón). Para acceder a las placas virtuales cada patólogo invitado dispuso de una cuenta personalizada de uso restringido y exclusivo.

¹ Sebastián Fernández Frez, "Trazabilidad de muestras en Anatomía Patológica Digital: aporte a la calidad", Actividad Formativa Equivalente para optar al grado de Magister en Informática Médica, Facultad de Medicina, Universidad de Chile, 2016.

² http://www.microscopiavirtual.cl/wp/?page_id=7. Accedido el: 08-12-2016

³ <http://www.nlhpc.cl/es/>. Accedido el: 08-12-2016

3.1.2 Creación de guías y formularios on-line para ingreso de resultados de evaluación visual

Con el propósito de familiarizar e informar a los patólogos con el método de trabajo on-line, se diseñó y elaboró una guía de usuario con instrucciones para el acceso a las placas virtuales, la realización de anotaciones y los procedimientos a seguir para la entrega de sus evaluaciones (ver Anexo III). El diagnóstico de cada placa virtual fue entregado por los patólogos a través de formularios on-line creados mediante las herramientas de Google Form (ver Anexo IV). Dichos formularios permiten el ingreso de datos y su posterior extracción bajo el formato de tablas excel.

3.1.3 Participación de patólogos invitados y realización de “pre-test”

Con respecto a la participación de patólogos, en base a datos existentes en la literatura relacionados con procesos de evaluación de biopsias [4, 7, 81], se determinó un número mínimo de 3 especialistas para la obtención de un diagnóstico referencial o *gold standard*.

Se invitó a participar a patólogos del sistema de salud público chileno, para realizar una evaluación visual de placas virtuales de biopsias procesadas con IHQ para los marcadores tumorales HER2, RE y RP utilizados en esta tesis. Los patólogos fueron informados que la evaluación se realizaría remotamente mediante internet, por tanto, no fue necesario su presencia física.

A través de un proceso preliminar, denominado “pre-test”, se familiarizaron de forma práctica con el flujo de trabajo, el acceso on-line a las placas virtuales, el uso de las herramientas de la plataforma de visualización y el ingreso de diagnóstico en los formularios on-line. El “pre-test” incluyó 6 casos para la evaluación del marcador HER2, 2 casos para RE y 2 para RP. El “pre-test” permitió estimar tiempos requeridos para el proceso de validación, identificando puntos críticos tales como el tiempo de respuesta por parte de los patólogos, la aceptación de la metodología de evaluación y se estableció además el número de ROIs a realizar por placa virtual.

Esta etapa también permitió definir las tareas específicas para los integrantes de CPDAI, así como detectar dificultades y proponer soluciones adaptadas a la realidad local, como fue el caso que presentó la extracción de ROIs desde la plataforma on-line (NDP.serve) a la plataforma NDP.analyze (Visiopharm). La metodología seguida de exportación de ROIs entre ambas plataformas se ilustra en la Figura 3.2.

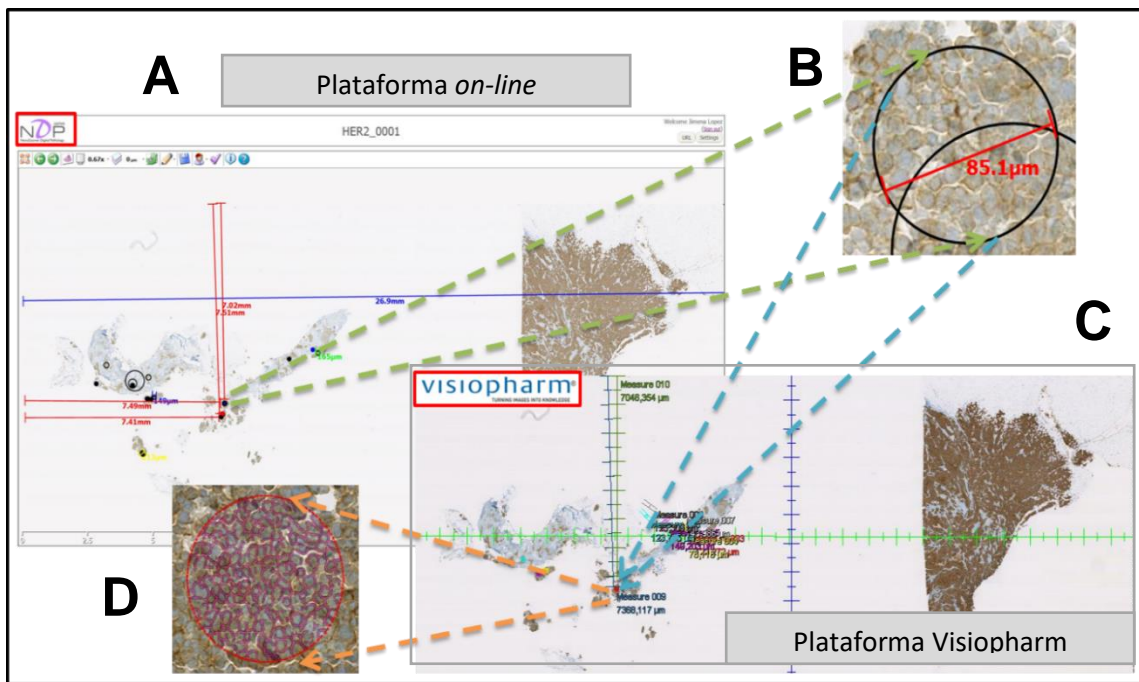


Figura 3.2: Proceso manual de exportación de ROIs desde la plataforma NDP.serve a la plataforma de Visiopharm. A) Imagen en pantalla de la plataforma NDP.serve, permite observar las placas virtuales en las que se han realizado las ROIs; B) Magnificación de una ROI con medición de su diámetro; C) Recreación de la ROI ilustrada en B en la plataforma Visiopharm, ubicada mediante comparación visual (pantallas paralelas); D) Magnificación de la ROI ilustrada en B una vez exportada y segmentada para la marcación de membrana de HER2 en la plataforma Visiopharm. Este proceso se iteró para cada una de las ROIs realizadas por los patólogos.

El diseño del flujo de trabajo resultante de la Etapa 1: Materiales y definición del flujo de trabajo, fue diagramado en notación BPMN (explicado en Sección 4.1; Ver además Anexo V). Una vez finalizado el pre-test, se procedió a implementar las restantes etapas del proceso de validación.

3.2 ETAPA 2: GENERACIÓN DE *GOLD STANDARD* POR PATÓLOGOS

El *gold standard* es la prueba considerada como la mejor alternativa diagnóstica existente y está ligada directamente a la evaluación visual de las placas virtuales realizada por los patólogos, los cuales entregaron ROIs con su respectivo diagnóstico y un diagnóstico general por placa, tanto para los casos de HER2 como para RE y RP. Los datos generados fueron utilizados tanto para configuración de los algoritmos como para la comparación de la cuantificación automatizada realizada.

Para la configuración y entrenamiento del algoritmo de cuantificación nuclear en receptores hormonales, fue necesaria la adición de un proceso de etiquetado de núcleos, como se mencionó en Sección 2.4.5. El algoritmo de cuantificación de membrana para HER2 no requiere ningún proceso de etiquetado de membrana, debido a que este ya que detecta específicamente la membrana marcada a diferencia del algoritmo de cuantificación nuclear que si requiere de entrenamiento para diferenciar núcleos positivos respecto a los negativos.

3.2.1 Obtención de ROIs y evaluación del estado de positividad de marcadores tumorales

Tres patólogos evaluaron las placas virtuales, el reporte de los resultados fue realizado de acuerdo a las guías de evaluación ASCO/CAP, para casos de HER2 se utilizó la clasificación 0/1+, 2+ y 3+, en tanto que para los de RE y RP se realizó de acuerdo al porcentaje de positividad nuclear. La evaluación consistió en la selección y dibujo de ROIs con su respectivo diagnóstico (anotación realizada en la plataforma de visualización) y un diagnóstico general por placa virtual (datos ingresados en los formularios; Anexo IV) lo que permitió constituir el marco referencial, o *gold standard* para la estandarización de los algoritmos. La Figura 3.3 muestra el proceso seguido por los patólogos para la evaluación de los tres marcadores utilizados.

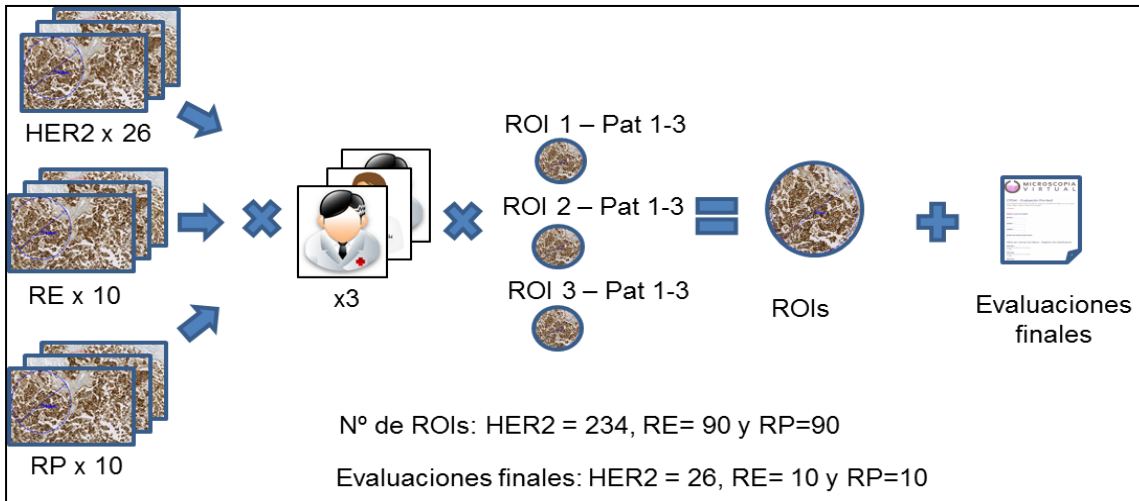


Figura 3.3: Proceso de evaluación de placas virtuales para los casos de HER2, RE y RP. A través de cuentas personalizadas, tres patólogos evaluaron las placas virtuales. De este proceso se obtuvieron en total 234 ROIs con su respectivo diagnóstico y un diagnóstico general por placa, todos los cuales fueron ingresados por los patólogos en los formularios on-line.

Dos ejemplos de las ROIs seleccionadas y dibujadas por dos patólogos, con su respectivo diagnóstico de evaluación para HER2 y RE se ilustran en la Figura 3.4

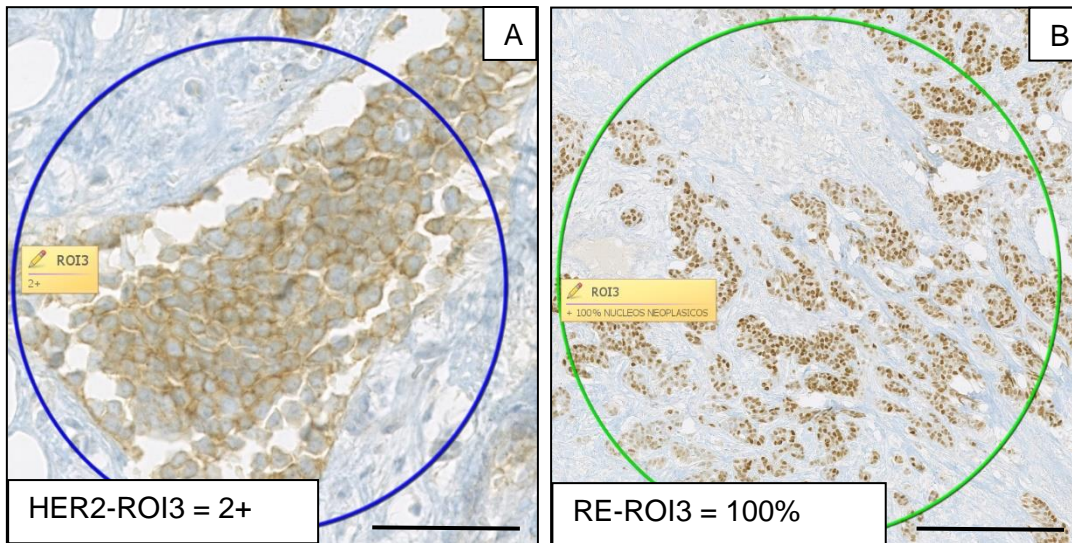


Figura 3.4: Ejemplo de ROIs seleccionadas y evaluadas por los patólogos. A) ROI de HER2 evaluada con diagnóstico 2+ (equivoco); B) ROI de RE con diagnóstico de 100% de positividad nuclear. Barra: A: 0,1 mm; B: 0,5 mm.

Las características de las ROIs seleccionadas, dibujadas y evaluadas por los patólogos fueron puestas bajo estudio. Se analizaron 180 ROIs provenientes de 20 placas virtuales de HER2. ROIs fueron sometidas a medición para conocer la variabilidad en términos de tamaño, localización (dispersión de ROIs) y clasificación (diagnóstico).

Para la detección de la variabilidad del tamaño en las ROIs se realizó la medición del área de cada una de ellas. Los resultados se presentan en la Sección 4.2.1.1

Para identificar áreas coincidentes y no-coincidentes y estudiar su localización se realizaron capturas de imagen macro de las placas virtuales con un tamaño de 1368x768 pixeles, como se ilustra en la Figura 3.5A. A estas imágenes se aplicaron matrices de 10x10 generando pequeñas sub-áreas, como se ilustra en la Figura 3.5B. Las ROIs ubicadas dentro de las sub-áreas fueron consideradas coincidentes y denominadas *hits*. Este proceso también permitió detectar la variabilidad existente en la clasificación de ROIs en áreas coincidentes. Los resultados de este análisis se presentan en la Sección 4.2.2.1

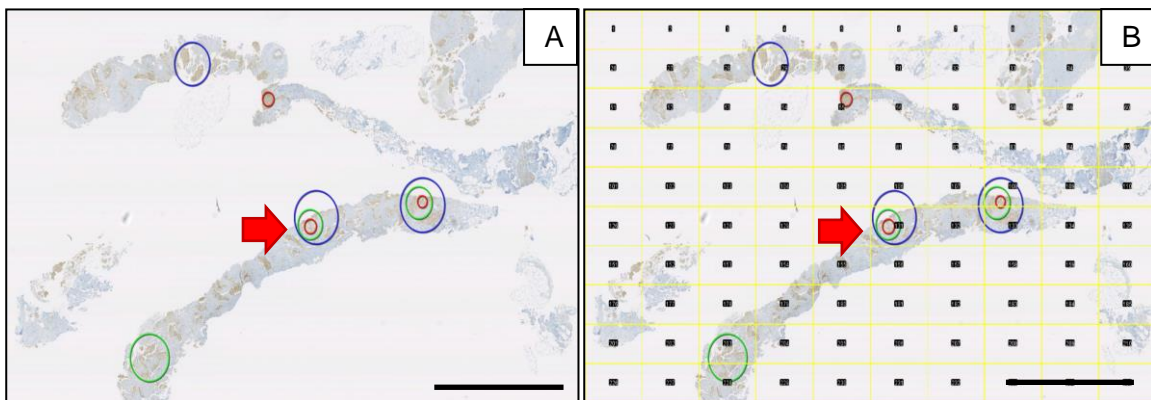


Figura 3.5: Matriz de localización de ROIs para establecer coincidencias entre patólogos. A) Placa virtual ilustrando las diferentes ROIs dibujadas por los patólogos según código de color (P1= verde, P2= azul y P3= rojo). Notar que en algunos casos hay coincidencia en la localización de las ROIs a pesar de las diferencias en tamaño; B) Matriz de 10x10 (líneas amarillas) que permite establecer áreas referenciales para las ROIs coincidentes y no-coincidentes. Barra: 2,5 mm.

3.2.2 Etiquetado de núcleos en receptores hormonales

Como se mencionó anteriormente, el algoritmo de cuantificación nuclear requiere de un proceso de entrenamiento para detección y diferenciación de núcleos positivos y negativos. Con este propósito se estableció un proceso adicional de etiquetado de núcleos. Cinco patólogos identificaron y etiquetaron manualmente cada núcleo en 3 ROIs representativas de RE seleccionadas del proceso mencionado en el punto anterior. De este proceso se obtuvieron 15 ROIs con núcleos categorizados y etiquetados bajo un código de colores (ver Anexo VI). En la Figura 3.6 se ilustra una ROI con núcleos etiquetados para el marcador RE y los resultados de la cuantificación de núcleos positivos y negativos.

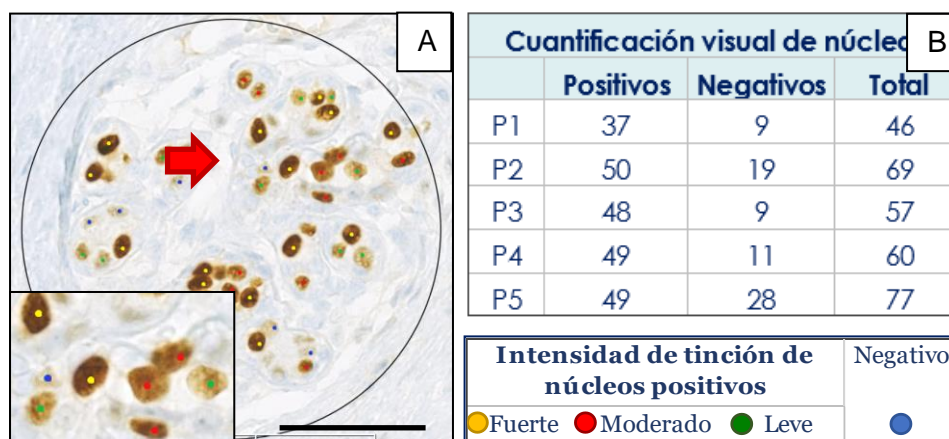


Figura 3.6: Identificación y etiquetado de núcleos positivos y negativos para RE. A) ROI-1 con núcleos categorizados manualmente según nivel de intensidad de tinción. Positivos: etiqueta amarilla = tinción fuerte, etiqueta roja = tinción moderada, etiqueta verde = tinción leve; Negativos = etiqueta azul; B) Tabla de resultados de la cuantificación manual del total de núcleos positivos y negativos realizado por cinco patólogos en la ROI mostrada en A. Notar la gran variación observada en la cuantificación de núcleos negativos (min 9; max 28) con respecto a los valores para núcleos positivos, que varían en menor grado. Barra: 0,25 mm.

3.3 ETAPA 3: ANÁLISIS DE IMAGEN

Esta etapa comprendió la configuración de los algoritmos y su aplicación posterior en la cuantificación de los marcadores tumorales en las mismas placas virtuales evaluadas por los patólogos. Con este propósito, los archivos de placas virtuales fueron importados a la base de datos del software de análisis de imagen NDP.analyze. El método de extracción y traspaso de ROIs, desde la plataforma

NDP.serve a la plataforma NDP.analyze, fue determinado en la etapa pre-test (Sección 3.1.3) y contempló una rutina de trabajo de dos pasos: i) la medición del diámetro de la ROI realizada por los patólogos y ii) la re-creación de dicha ROI en la plataforma NDP.analyze. Se realizó a través de una comparación visual de la ubicación del ROI en cada placa virtual (Figura 3.2). Este método se aplicó para todos los casos de HER2, RE y RP. Una vez que las ROIs fueron establecidas en la plataforma NDP.analyze, se procedió a configurar los algoritmos para cada cuantificación.

Como se mencionó en la Sección 2.5, NDP.analyze cuenta con tres algoritmos básicos pre-definidos que permiten la detección de marcadores biológicos localizados en: i) membrana celular, ii) núcleo y iii) citoplasma. Los algoritmos de análisis de imagen para membrana celular y núcleo fueron configurados para la cuantificación de los marcadores utilizados en la presente tesis. A continuación, se describe el abordaje para configurar los algoritmos tanto para la cuantificación del marcador de membrana HER 2 como para los marcadores nucleares RE y RP.

3.3.1 Estandarización del algoritmo conectividad de membrana

Para propósitos de configuración de parámetros del algoritmo se seleccionaron 12 en las que hubo concordancia en su evaluación general, con un total de 108 ROIS evaluadas y/o clasificadas, las que fueron utilizadas como conjunto de entrenamiento. Para la etapa posterior de validación, se consideraron 14 placas virtuales restantes, las que proporcionaron un total de 117 ROIs como conjunto de validación.

Como se mencionó en Sección 2.4.3.1, este algoritmo requiere de tres parámetros: la ROI, la sensibilidad y la conectividad. Las ROIs fueron provistas por los patólogos y constituyeron las áreas en las cuales se realizó la cuantificación. Los valores de los parámetros de sensibilidad y conectividad descritos en la Sección 2.5.1, fueron definidos en esta etapa.

Debido a que se requiere buscar un rango de sensibilidad que permita obtener una mayor separación de las clases 0/1+, 2+ y 3+, se realizaron pruebas preliminares configurándose el algoritmo en el rango total de sensibilidad entre 0 y 100%, con intervalos de 5%. Con esta configuración se segmentaron 20 ROIs obtenidas del “pre-test”. Los valores de conectividad obtenidos del análisis de las ROIs dentro este rango fueron representados mediante un gráfico de curvas ilustrado en la Figura 3.7, en la que claramente se evidencia una fluctuación de los valores de conectividad respecto a cada nivel de sensibilidad.

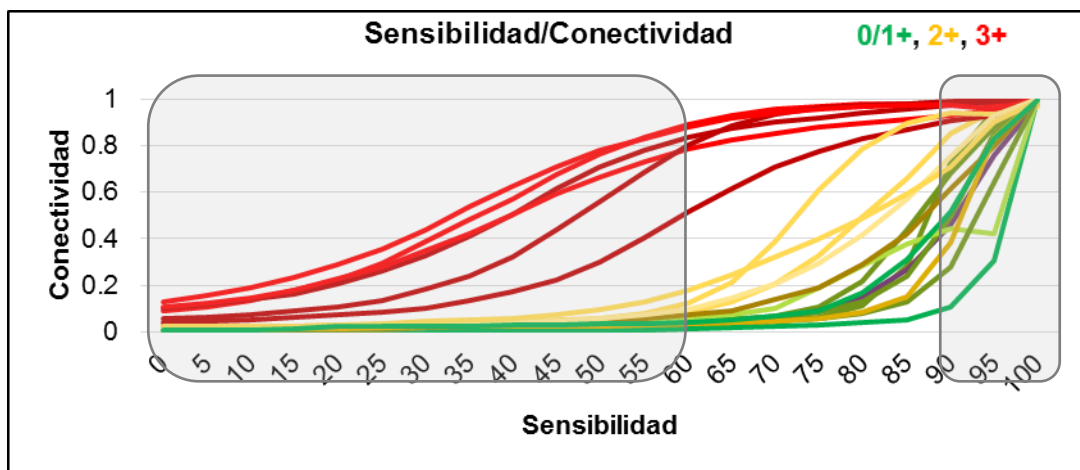


Figura 3.7: Configuración del parámetro Sensibilidad para la obtención de los valores de conectividad para HER2. Los valores de conectividad observados en las ROIs fueron agrupados bajo un código de colores según la clasificación ASCO/CAP (0/1+, verde; 2+, amarillo; 3+ rojo) obtenidas en un rango de sensibilidad de 0% a 100%. Se observa la fluctuación de con intervalos menores a 60% y mayores a 90% de sensibilidad los valores de conectividad no representan de valores de conectividad, representando características reales de la membrana marcada. Se observa además que las curvas tienen la tendencia de inclinarse hacia 0 (< 60%) o 1 (> 90%).

Este análisis mostró que sensibilidades menores a 60% y mayores a 90%, los valores de conectividad de las ROIs analizadas, tienden a ser constantes y poco variables, como es ilustrado en la Figura 3.8.

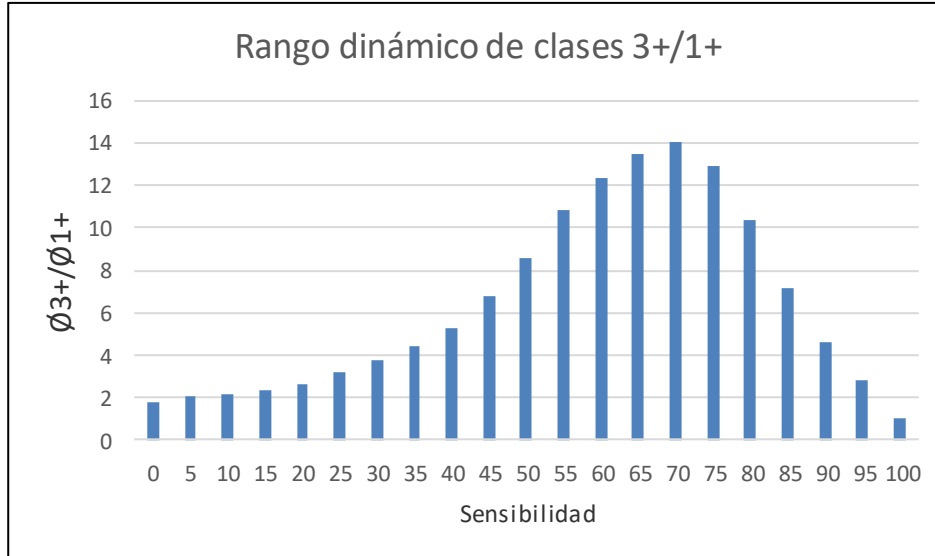


Figura 3.8: Coeficientes de fluctuación de grupo 3+ respecto a 1+ en función de la sensibilidad. Se observa que en el rango de 55% a 90% de sensibilidad los valores de conectividad del grupo 3+ respecto a 0/1+ presentan una mayor separación de clases, en tanto que valores menores o a 55% o mayores a 90% los valores de conectividad tienen una tendencia a ser constantes.

Los resultados de este análisis permitieron determinar y/o configurar el algoritmo en el rango de 60-90% de sensibilidad con intervalos de 5%. Dentro de este intervalo de sensibilidad se obtuvieron valores de conectividad tanto para el conjunto de entrenamiento como el de validación. En la Figura 3.8 se observa el gráfico de curvas de los valores de conectividad obtenidas del análisis de ROIs dentro del rango ya definido.

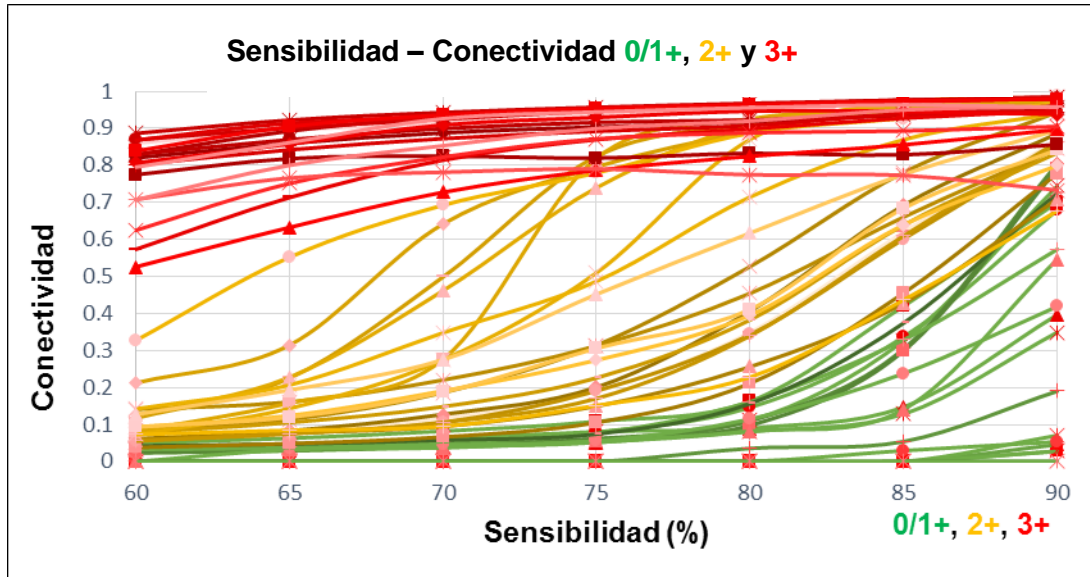


Figura 3.9: Relación sensibilidad – conectividad. Los valores de conectividad de ROIs en estudio fueron sometidos al análisis con el algoritmo en el rango de sensibilidad 60-90%. Los valores de conectividad de ROIs 0/1+ (verde) y 3+ (rojo) muestran una evidente separación. Por el contrario, los valores de conectividad de ROIs clasificados 2+ (amarillo) tienen una tendencia a sobrelaparse (*overlapping*) respecto a la clasificación 0/1+. Se observa que a 75% de sensibilidad la separación de los valores de conectividad para cada grupo es mucho más evidente.

Para la traducción de los valores numéricos de conectividad a las categorías de clasificación ASCO/CAP (0/1+, 2+ y 3+), se contó con dos opciones: (i) trabajar con valores de conectividad obtenidos en un único valor de sensibilidad (un valor de sensibilidad que visualice una mayor separación de grupos y definir puntos de corte), o ii) trabajar con todos los valores de conectividad obtenidos en el rango de 60-90% de sensibilidad. El desarrollo de ambas metodologías se describe en la siguiente a continuación.

3.3.2 Traducción de valores de conectividad a categorías ASCO/CAP

Los valores de conectividad fueron traducidos a las categorías ASCO/CAP (0,1+, 2+ y 3+) a través de clasificadores de máquinas de aprendizaje, mediante dos métodos: i) el primero mediante clasificadores de árboles de decisión CART, alineado al método propuesto por los desarrolladores del algoritmo (Visiopharm) utiliza los valores de conectividad obtenidos a un 75% de sensibilidad con el fin de

definir valores de corte y su asignación o mapeo a la clasificación ASCO/CAP (se realizaron pruebas con el 70% y 80% de sensibilidad, los resultados del entrenamiento del clasificador fueron bajos, ver Anexo VII) y ii) utilizando el clasificador SMO de *Support Vector Machine* para el cual se utilizaron todos los valores de conectividad obtenidos en el rango 60-90% de sensibilidad. El método SMO fue seleccionado de acuerdo al tamaño de la muestra estudiada (N relativamente pequeño para uso de otro clasificador), particularidades del clasificador y su uso en estudios realizados en cáncer, como fue mencionado en la Sección 2.5.2. En ambas metodologías, para el aprendizaje de los clasificadores, se utilizaron los datos del conjunto de entrenamiento, esto es, valores de conectividad de ROIs asociadas a sus respectivas clasificaciones. La secuencia del proceso de aprendizaje se ilustra en la Figura 3.9, en tanto que información detallada de ambos métodos es provista en la Sección 2.5.

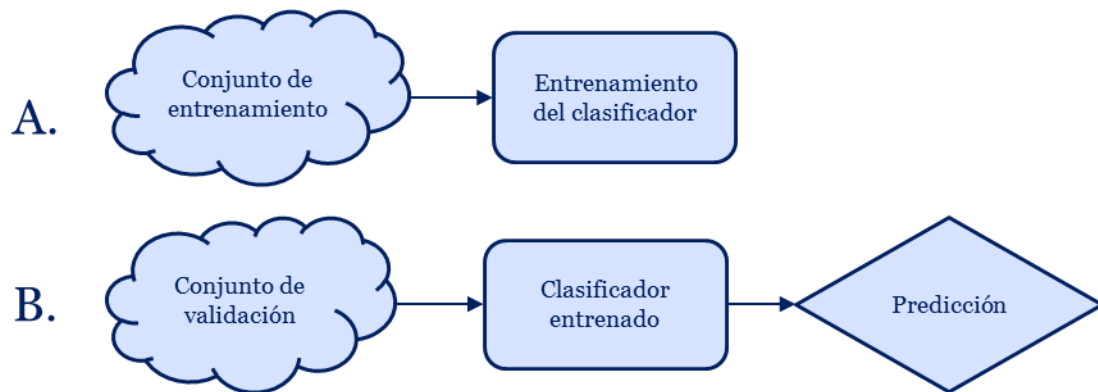


Figura 3.10: Proceso de aprendizaje del clasificador (algoritmo) a partir de datos del conjunto de entrenamiento. A) Los datos de entrada corresponden a valores de conectividad de 108 ROIs etiquetadas según clasificación provista por los patólogos. El algoritmo de máquinas de aprendizaje va aprendiendo a partir de la información que provee el conjunto de entrenamiento, según el criterio utilizado por los patólogos (diagnóstico de las ROIs); B) El clasificador entrenado recibe los valores de conectividad, del conjunto de validación y procede a clasificar los valores de conectividad y entregar una clasificación (0/1+, 2+ y 3+).

Los rangos de conectividad asociados a las distintas categorías ASCO/CAP con la metodología CART a 75% de sensibilidad se presentan en la Sección 4.3.1. A partir de la clasificación de las ROIs se obtuvo una clasificación general para caso de HER2 (ver Anexo VIII). Los puntos de corte obtenidos mediante este método fueron aplicados en el conjunto de validación (117 ROIs). Respecto al clasificador SMO,

este fue entrenado con todos los valores de conectividad obtenidos en el rango de 60 a 90% de sensibilidad. Posteriormente se utilizó el clasificador SMO entrenado para traducir los valores de conectividad del conjunto de validación (ver Anexo IX). Los resultados de ambas metodologías se presentan en el Capítulo 4, Sección 4.3.

3.3.3 Estandarización de algoritmos de cuantificación nuclear

Como se describió en la Sección 2.4.3.2, la configuración del algoritmo de cuantificación nuclear requirió de una etapa de entrenamiento para diferenciación de núcleos positivos respecto a los negativos. En este proceso se ajustaron diversos parámetros, tales como, sensibilidad, forma del núcleo, color e intensidad de tinción.

Posterior al proceso de entrenamiento se siguió la metodología de configuración para obtención de porcentajes de positividad. Sin embargo, se presentaron dificultades para obtener los porcentajes de positividad, debido a que las muestras presentaban una tinción tenue e inespecífica en los núcleos negativos, similar a la presentada al tejido de fondo ocasionando una incorrecta detección por parte del algoritmo. Para solucionar el problema, se propuso una nueva metodología para la segmentación y cuantificación de núcleos positivos. Ambas metodologías se describen con detalle a continuación.

3.3.4 Configuración del algoritmo Visiopharm

Para la configuración de este algoritmo se proporcionaron características respecto de la intensidad de tinción, textura y tamaño del núcleo, a partir de los ejemplos suministrados por los patólogos. Tales características fueron definidas con las herramientas de análisis de imagen del algoritmo predefinido denominado "*Nuclei App*". Se asignaron etiquetas como se observa en la Figura 3.10, las que permitieron dar instrucciones al algoritmo para la detección de núcleos tanto positivos como negativos. Con el propósito de precisar dicha detección como también los niveles de intensidad de la marcación nuclear, se ajustaron valores de sensibilidad.

Con la metodología, el algoritmo permitió la correcta identificación y cuantificación de núcleos positivos comparados con los núcleos etiquetados y cuantificados por los patólogos. En la Figura 3.11A se ilustra la detección de núcleos positivos obtenida mediante el algoritmo “*Nuclei App*” [57].

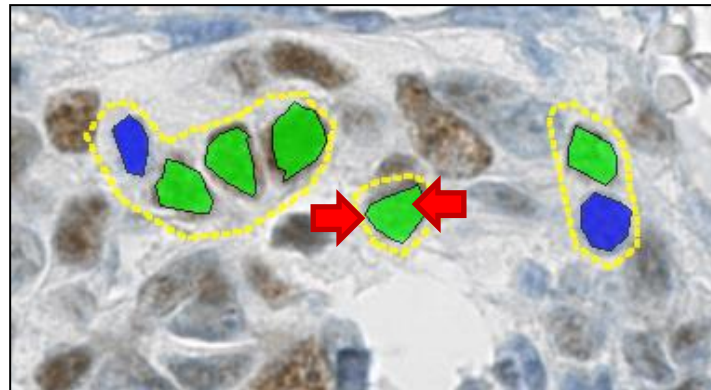


Figura 3.11: Etiquetado de núcleos para entrenamiento del algoritmo de cuantificación. Se ilustra un área de análisis (zonas de entrenamiento) con presencia de núcleos positivos y negativos. Se observa la asignación de etiquetas de color verde a algunos núcleos positivos (flecha roja) y de color azul a los negativos, presentes en ROIs delimitadas por un contorno amarillo. Barra: 0,1mm.

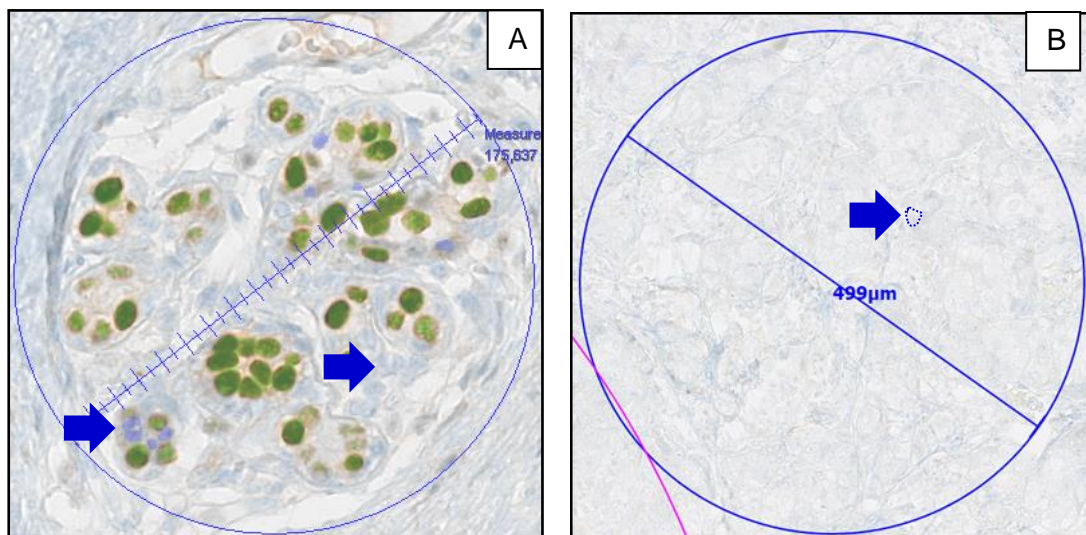


Figura 3.12: Detección de núcleos mediante el algoritmo Visiopharm Nuclei App. A) ROI mostrando detección de núcleos positivos (etiqueta verde) y negativos (etiqueta azul; flecha azul) excluyendo zonas de estroma y tejido de fondo. B) Imagen representativa de ROI seleccionada por el patólogo, considerada con diagnóstico de 0% de positividad, donde se puede apreciar la tinción inespecífica de núcleos negativos, difíciles de diferenciar del tejido de fondo (línea de puntos). Se ilustra el diámetro de cada ROI.

Con respecto a los núcleos negativos, el algoritmo no permitió detectarlos de manera confiable pues en nuestras muestras, en particular, estos presentaron una contra-tinción muy tenue e inespecífica, similar a la presentada por el estroma o tejido de fondo, como se observa en la Figura 3.11B . En consecuencia, no fue posible obtener el número total de núcleos, es decir, la suma de núcleos marcados (positivos para RE, color café) y no marcados (negativos para RE), siendo este valor indispensable para establecer el porcentaje de positividad de la muestra.

Esta dificultad fue resuelta mediante la creación y configuración de un nuevo algoritmo de metodología propia, el que fue denominado *Region Growing* (RG), para lo cual se utilizaron las herramientas de pre y post procesamiento de análisis de imágenes que proporciona el software NDP.analyze que, a diferencia del algoritmo Visiopharm, no requiere la cuantificación de núcleos negativos.

3.3.5 Configuración algoritmo propuesto “*Region Growing*”

RG fue configurado específicamente para la detección y cuantificación de núcleos positivos. La detección se realizó a partir de la expansión progresiva de un área que va discriminando al mismo tiempo zonas en las que los núcleos positivos están ausentes, es decir, excluye núcleos negativos, estroma y/o tejido de fondo. Este algoritmo crea una región delimitada por un contorno o barrera de exclusión que segmenta sólo núcleos positivos. El cálculo del porcentaje de positividad se presenta en la Sección 3.3.6. A continuación, se describen los pasos de configuración del algoritmo RG:

1. Identificar núcleos positivos en base a la morfología e intensidad de tinción (proceso de entrenamiento similar al mostrado en la Figura 3.12A solo para núcleos positivos). Se ajustan niveles de sensibilidad para lograr la mejor detección posible.
2. Establecer una condición de detección en la cual se analiza la presencia de áreas con acumulación de núcleos positivos (zonas cancerígenas). El algoritmo considera como núcleo a áreas $>5\mu\text{m}^2$, etiquetándolas como tal y crea una ROI redefinida que se va expandiendo hasta cubrir todos los

núcleos positivos dentro de una vecindad, como se ilustra en la Figura 3.12B . Esta nueva ROI debe detectar un área $>250\mu\text{m}^2$ cubierto por núcleos positivos para dar cuenta de la agrupación de estos.

- Realizar una nueva detección de los núcleos positivos dentro de la ROI redefinida (Figura 3.12C).
- Calcular el área cubierta de la ROI redefinida ($A_{roi\ pos}$, líneas rosadas), y el área total de núcleos positivos marcados ($A_{n\ pos}$, etiqueta verde).
- Calcular el coeficiente de positividad (cp_{pos}) se obtiene a partir del coeficiente entre las áreas de núcleos positivos marcados y detectados por el algoritmo (etiqueta verde) y el área de la ROI redefinida (línea rosadas) como se ilustra en la Ecuación 1:

$$cp_{pos} = \frac{A_{n\ pos}}{A_{roi\ pos}} \quad [1]$$

- Ajustar los niveles de sensibilidad para detección adecuada, siguiendo la información de las ROIs etiquetadas provistas por los patólogos (Sección 3.1.1.1).

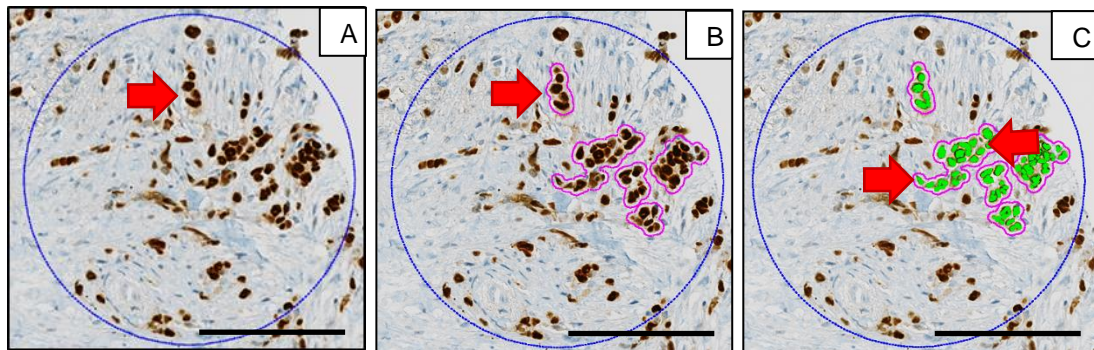


Figura 3.13: Secuencia de detección específica de núcleos positivos mediante el algoritmo RG. A) ROI con núcleos positivos para RE. B) Detección mediante algoritmo RG de áreas dentro la ROI con núcleos positivos (delimitadas por un contorno rosado). A medida que el algoritmo detecta núcleos positivos en las vecindades, las áreas se van expandiendo. C) Para precisar dicha detección, el algoritmo GR itera la detección en las áreas expandidas etiquetando los núcleos positivos (etiqueta verde). Barra: 0,25 mm.

La segmentación y cuantificación de núcleos mediante el algoritmo RG permitió obtener un coeficiente de positividad (cp_{pos}). El cálculo del porcentaje de positividad se describe a continuación.

3.3.6 Metodología de cálculo del porcentaje de positividad

La metodología del cálculo del porcentaje de positividad se realizó usando un valor normalizado al 100%, para este propósito se utiliza como referencia la muestra control, incluida siempre en la placa. Dado que esta tiene 100% de positividad, es posible compararla con la muestra en estudio y obtener de esta manera el porcentaje de positividad. En la parte derecha inferior de la Figura 3.13A, se observa un portaobjeto representativo con una muestra control y la muestra a ser evaluada.

Para obtener el valor del coeficiente de control positivo tanto para casos de RE (10 casos) y RP (10 casos) se seleccionaron y dibujaron 10 ROIs con un tamaño estandarizado (diámetro = 1000 μm) en muestras de control positivo de cada placa virtual como se observa en la Figura 3.13A. Posteriormente estas se segmentaron y cuantificaron mediante el algoritmo RG, obteniendo el coeficiente de positividad de la muestra control ($cp_{cp_{pos}}$). La Figura 3.13B ilustra la segmentación de una ROI, denotando el área de la ROI redefinida expandida (línea magenta) y agrupando núcleos positivos (áreas verdes), se observa además áreas discriminadas en las cuales los núcleos positivos se encuentran ausentes (en la parte inferior derecha se visualiza el valor del coeficiente de positividad resultante).

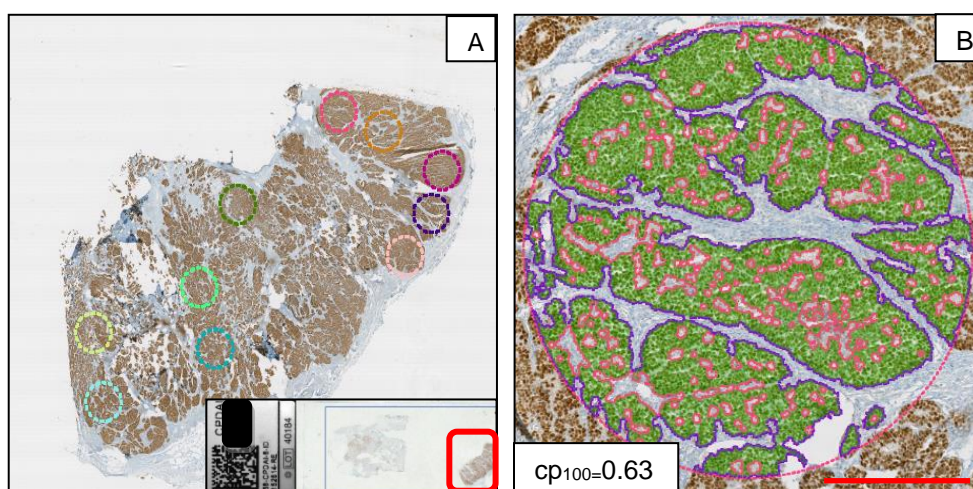


Figura 3.14: Imagen representativa de un control positivo en una placa. A) Muestra control ilustrando 10 ROIs en diferentes colores con 100% de positividad nuclear. El algoritmo RG permitió obtener los respectivos coeficientes de positividad de la muestra control que se utilizaron para el cálculo del porcentaje de positividad. En la parte inferior derecha se ilustra el portaobjeto completo, con el tejido de la muestra a estudiar a la izquierda y delimitado en rojo a la derecha se aprecia el control positivo; B) ROI de la muestra control segmentada por el algoritmo RG. En la parte inferior izquierda se muestra el valor del coeficiente de positividad, en este caso 0.63. Barra de calibración: 0.5 mm

Los coeficientes control positivo $cp_{cp_{pos}}$, obtenidos, se promedian para obtener un solo coeficiente de control positivo por placa, como se representa en la Ecuación 2:

$$cp_{100} = \frac{1}{n} \sum_{i=1}^n cp_{cp_{pos}(i)} \quad [2]$$

cp_{100} = Coeficiente de control positivo por placa
 $cp_{cp_{pos}}$ = Coeficiente de positividad de la muestra control
 n = Número de rois control por placa ($n = 10$)

Del análisis de ROIs provistas por los patólogos mediante el algoritmo “RG” se obtuvieron los coeficientes de positividad cp_{pos} tanto para casos de RE y RP . Se calcularon los porcentajes de núcleos positivos np_{pos} en base de la razón entre coeficientes de positividad obtenidos de las ROIs en estudio cp_{pos} y el coeficiente de positividad obtenidos en los controles cp_{100} representados en la Ecuación 3:

$$np_{pos} = \frac{cp_{pos} \times 100}{cp_{100}} \quad [3]$$

np_{pos} = Porcentaje de positividad normalizado
 cp = Coeficiente de positividad en muestras

Los coeficientes de positividad de la muestra control positivo tanto para casos RE como para RP fueron analizados estadísticamente mediante el test de normalidad de Shapiro-Wilk y pruebas paramétricas Tukey's HSD (*Honest significant difference*) paramétrico y no paramétrico de Kolmogórov-Smirnov previo a su aplicación y uso en la cuantificación de ROIs provistas por los patólogos.

3.4 ETAPA 4: ANÁLISIS ESTADÍSTICO DE CONCORDANCIA PATÓLOGOS VS ALGORITMOS

Los resultados obtenidos en la etapa anterior fueron analizados estadísticamente a través de los paquetes de software estadísticos IBM SPSS (*Statistical Package for the Social Sciences*)¹ y R (*Project for Statistical Computing*)², con el propósito de determinar la concordancia de los resultados obtenidos mediante los algoritmos de cuantificación y los resultados entregados por los patólogos. Además, se realizaron cálculos adicionales para determinar la normalidad de los datos.

Para la traducción de valores de conectividad de HER2 a la clasificación ASCO/CAP se utilizaron las plataformas WEKA v.3.6 (*Waikato Environment for Knowledge Analysis*)³ y MATLAB v.2015 (MATrix LABORatory)⁴. La primera es una plataforma para aprendizaje automático y minería de datos, desarrollada por la Universidad de Waikato en Nueva Zelanda; la segunda es una plataforma de *software* matemático que ofrece un entorno de desarrollo integrado, con lenguaje de programación propio.

Los resultados del análisis se presentan en el Capítulo 4

¹ <http://www.ibm.com/analytics/us/en/technology/spss/>. Accedido el: 15-12-2015

² <https://www.r-project.org/>. Accedido el: 15-02-2016

³ <http://www.cs.waikato.ac.nz/ml/weka/>. Accedido el: 15-12-2015

⁴ <https://www.mathworks.com/products/matlab.html>. Accedido el: 15-12-2015

4.RESULTADOS

4.1 IMPLEMENTACIÓN

Los resultados del objetivo específico número uno, Etapa 1: Materiales y definición del flujo de trabajo (ver Capítulo 3, Sección 3.1), permitieron establecer los requerimientos para la implementación del proceso de validación de PD en CM. En esta etapa de implementación se establecieron las tareas y actividades específicas, el personal responsable, así como la secuencia a seguir dentro del flujo de trabajo global del proceso. Las diversas tareas y/o actividades a llevar a cabo en: i) el Laboratorio de Anatomía Patológica del HCUCH (institución externa), ii) CPDAI y iii) patólogos invitados, fueron representadas en un diagrama en notación estándar de procesos BPMN (definida en Capítulo 2, Sección 2.6). El diagrama de flujo del trabajo global, así como las tareas específicas correspondientes a cada una de las tres instancias mencionadas se encuentran detalladas en el Anexo V.

Con respecto a las responsabilidades asignadas, la institución externa seleccionó los casos y entregó a CPDAI las placas a ser digitalizadas. Las tareas del personal CPDAI consistieron en: i) la gestión del flujo de trabajo y ii) la configuración y estandarización de los algoritmos de cuantificación. Por otra parte, la actividad realizada por los patólogos invitados consistió en la selección de ROIs y generación de un *gold standard*.

Un ejemplo de la diagramación de una actividad es ilustrado en la Figura 4.1, donde se representa el flujo de trabajo a través del cual los patólogos establecieron las ROIs representativas de zonas de tumor invasivo para cada placa. Tal como se describió en el Capítulo 3, Sección 3.2.1, esta tarea también incluía la entrega de evaluaciones generales de cada placa analizada. El flujo de trabajo ilustrado en la Figura. 4.1 fue utilizado para la evaluación del estado de positividad de los marcadores tumorales HER2, RE y RP.

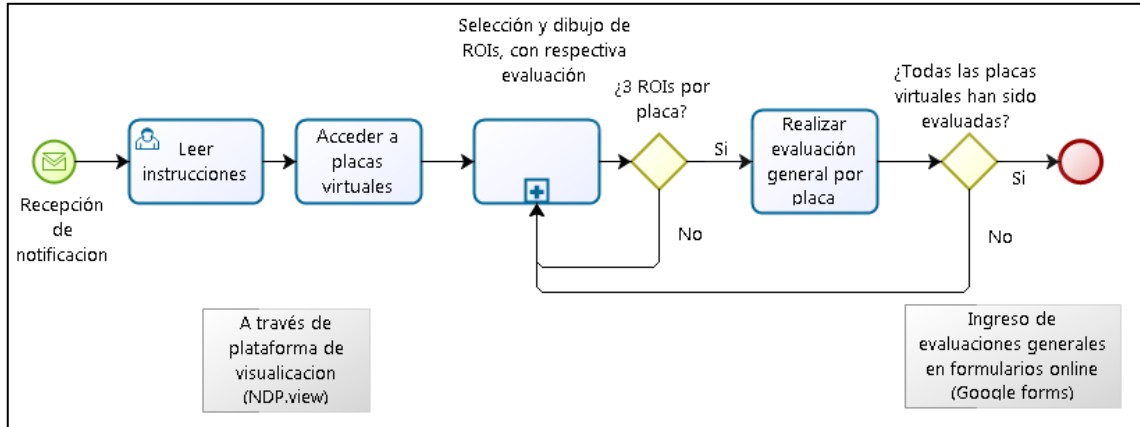


Figura 4.1: Flujo de trabajo seguido por los patólogos. Se ilustra el proceso de selección de ROIs y su evaluación IHQ. El flujo de trabajo se inicia con la recepción de la notificación (círculo verde) enviada por CPDAI a cada patólogo. Las tareas son identificadas en bloques (rectángulos celestes) donde la condición binaria (representada por un rombo amarillo) permite identificar si la tarea está realizada o se requiere una nueva iteración hasta finalizar el proceso (círculo rojo).

4.2 EVALUACIÓN VISUAL DE PLACAS VIRTUALES – GOLD STANDARD

Los resultados obtenidos de la generación del *gold standard*, tanto para HER2 como para los receptores hormonales RE y RP fueron puestos bajo estudio a fin de analizar las ROIs y los resultados de la evaluación realizada por los patólogos.

4.2.1 Análisis de datos para HER2

Se obtuvo un total de 234 ROIs diagnosticadas con sus respectivas 78 clasificaciones generales, como producto de la evaluación de 26 placas virtuales por 3 patólogos. El análisis de los datos permitió verificar la existencia de variabilidad en: i) el tamaño de las ROIs, ii) localización de ROIs y iii) clasificación del estado de positividad de HER2 tanto por ROIs coincidentes en una misma área, así como también por placa virtual. A continuación se describe en detalle los resultados de este análisis.

4.2.1.1 Variabilidad en el tamaño de las ROIs

El establecimiento de ROIs por los patólogos en cada una de las placas virtuales, evidenció que existen importantes diferencias en los tamaños de dichas regiones. Los resultados de la medición de sus áreas mostraron que estas fluctúan entre los 18 y 356 mm², tal como se ilustra en la Figura 4.2. Esta variación en el tamaño de las ROIs corresponde aproximadamente a una relación 1:20 y cada área agrupó entre 50 a 1000 células, respectivamente, población sobre la cual se realizó la evaluación de la expresión de HER2, tanto visualmente por los patólogos como en forma automatizada mediante el algoritmo ‘conectividad de membrana’.

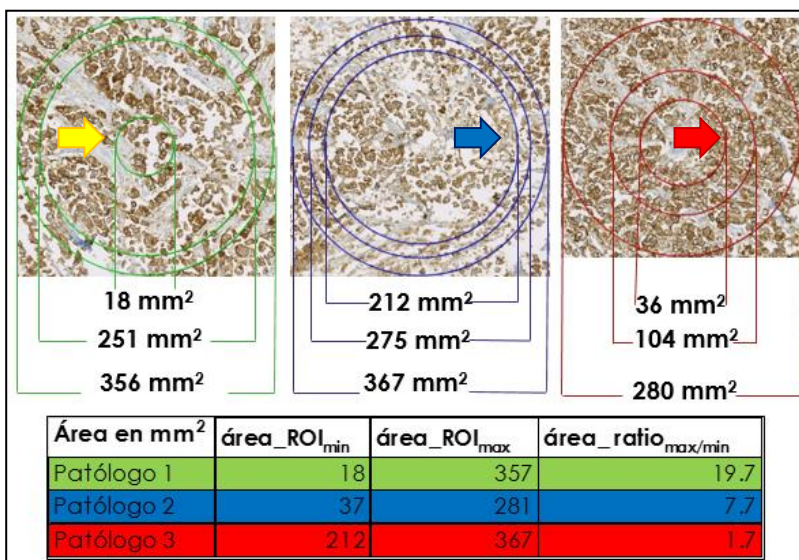


Figura 4.2: Variabilidad en el tamaño de las ROIs dibujada por los patólogos. El panel superior ilustra las áreas dibujadas por tres patólogos (verde, azul, rojo). Algunos patólogos dibujaron ROIs de áreas muy variables (verde y rojo, con una relación de 18:356 y 36:280 mm², respectivamente). En el panel inferior se muestra el valor de las áreas mínimas, máximas y su relación área-ratio, evidenciando la gran variabilidad observada.

4.2.1.2 Variabilidad en la localización de ROIs

Como se describió en el Capítulo 3 Sección 3.2.1, se utilizó una matriz para precisar las áreas de coincidencia y no-coincidencia, con el fin de analizar la localización de las ROIs presentes y definir las en términos de “hits”. Los resultados obtenidos fueron categorizados de la siguiente manera: i) áreas coincidentes entre

los tres patólogos, fueron denominadas 3 *hits*, ii) entre dos patólogos 2 *hits* y por último iii) áreas donde no hubo coincidencia fueron denominadas 1 *hit* (e. g., ROIs individuales no coincidentes). Esta categorización reveló que existe una gran predominancia de áreas de no-coincidencia, poniendo de este modo en evidencia la existencia de una considerable variabilidad en términos de localización de las ROIs, lo que sugeriría que cada patólogo selecciona ROIs. En la Figura 4.3 se ilustra un gráfico de barras con el número de hits detectados en las placas bajo estudio.

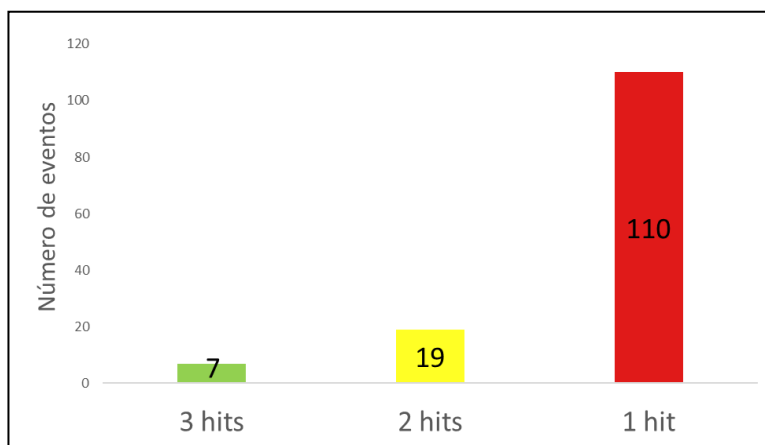


Figura 4.3: Numero de coincidencia entre la localización de ROIs. La presencia de áreas de no coincidencia entre los patólogos es el evento más frecuente (barra roja). La segunda frecuencia corresponde a áreas de coincidencia entre dos patólogos (barra amarilla). Los casos menos frecuentes, en los que los tres patólogos coincidieron, solo corresponden a 7 áreas (barra verde).

En el siguiente punto se describe la variabilidad encontrada en el diagnóstico de ROIs para HER2 presente en áreas de 3 y 2 *hits*.

4.2.1.3 Variabilidad en la clasificación de HER2

El resultado del análisis de las evaluaciones de las ROIs presentes en áreas de coincidencia de 3 y 2 *hits* evidenció además la existencia de variabilidad en la clasificación entregada. Aproximadamente un 15% de estas ROIs presentaron variabilidad, con su clasificación oscilando entre los estados 0/1+ y 2+ (Figura 4.4), que son los que presentan mayor grado de dificultad al momento de su evaluación.

En el restante 86% se encontró concordancia. Cabe señalar que el porcentaje de variabilidad observado en esta tesis es cercano a lo descrito en la literatura, donde se reporta que un 20% de las biopsias analizadas arrojan variabilidad inter-observador en su diagnóstico (referencias). Los casos en los cuales se observó variabilidad fueron en casos 0/1+ y 2+ y no así casos 0/1+ y 3+.

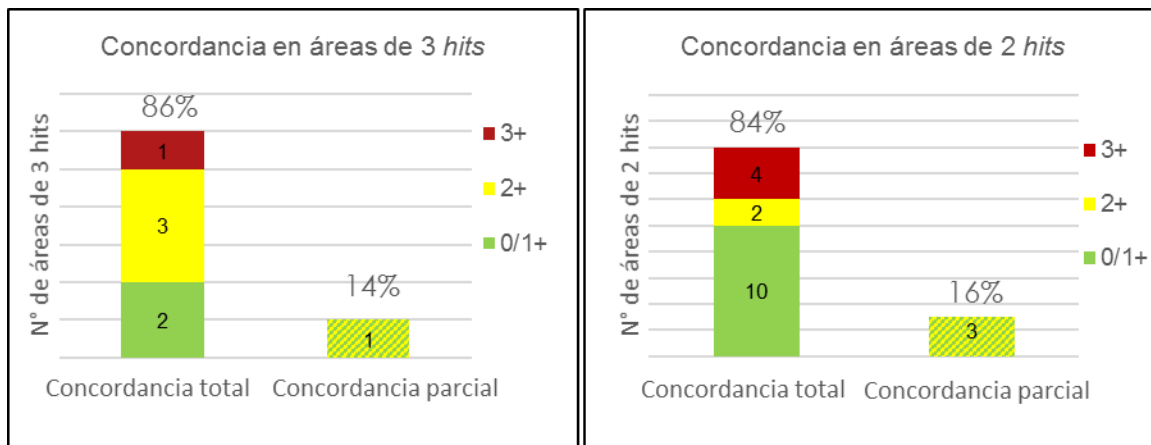


Figura 4.4: Variabilidad en la clasificación de áreas de 3 y 2 hits. A) Clasificación de ROIs en 7 áreas de 3 hits (coincidencia entre 3 patólogos), se observa una variabilidad del 14 % en casos con clasificaciones 0/1+ y 2+ en una misma área (concordancia parcial); B) Concordancia observada en la clasificación de 19 áreas de 2 hits (coincidencia entre dos patólogos), con una variabilidad observada del 16% en casos 0/1+ y 2+.

El análisis de ROIs según *hits* permitió además observar particularidades de un área de análisis. Un área de 3 *hits* presentó una ROI evaluada como carcinoma ductal *in situ* (CDIS, Figura 4.5A), en cuyo caso el patólogo sugirió contar con una placa adicional de un corte de la misma muestra procesada con tinción de hematoxilina-eosina, para confirmar el diagnóstico.

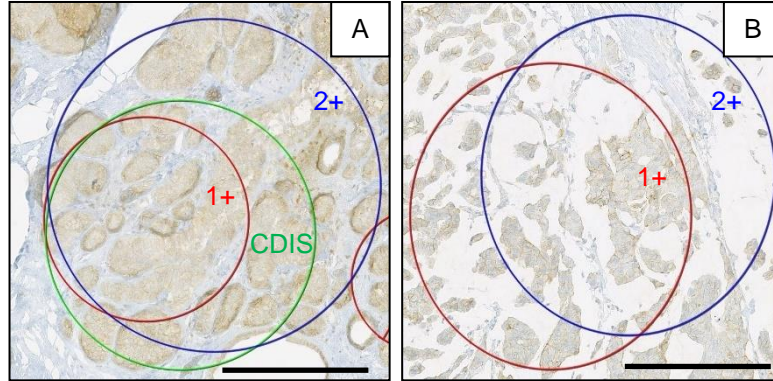


Figura 4.5: Ejemplo de variabilidad en la clasificación de ROIs en áreas de 3 y 2 hits. A) ROIs de un área de 3 hits clasificada como 1+, CDIS y 2+, es decir, con diagnósticos diferentes entregados por los tres patólogos. B) ROIs de un área de 2 hits con clasificaciones 1+ y 2+, para una evaluación realizada en áreas relativamente cercanas. Barra 0.25 mm.

También se analizó el porcentaje de concordancia entre los patólogos con respecto a las evaluaciones generales por placa virtual. Un 46% de las 26 placas utilizadas presentaron variabilidad, la que se distribuyó en las cuatro categorías tradicionales: 0/1+, 2+ y 3+, correspondiendo a 12 de los 26 casos analizados. Agrupando las dos primeras categorías (0+ y 1+), ambas consideradas como diagnóstico negativo, se observó una concordancia del 69%, la que corresponde a 18 de los 26 casos. En los 8 casos restantes se observó una concordancia parcial, es decir, dos de los tres patólogos concordaron en su clasificación. Esta concordancia parcial fue observada en casos con clasificaciones 0/1+ y 2+ (Figura 4.6).

Estos resultados se pueden interpretar bajo la premisa que cada patólogo aplica su propio criterio de selección de ROIs para la evaluación general de cada caso.

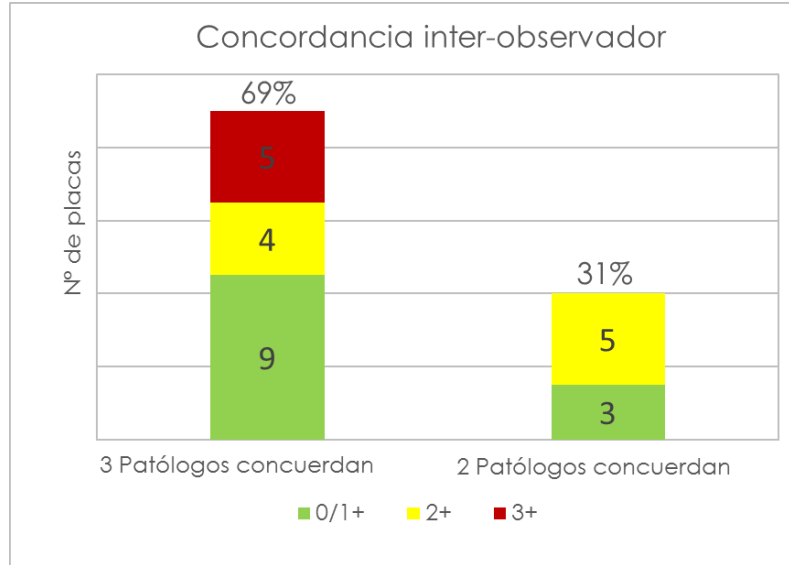


Figura 4.6: Concordancia y variabilidad inter-observador. La barra de la izquierda corresponde a la concordancia en la clasificación de HER2 por tres patólogos, en tanto que a la derecha se muestra la concordancia entre solo dos patólogos. Los casos con concordancia parcial o variabilidad detectada (31%) corresponden a clasificaciones 0/1+ y 2+ que presentan mayor dificultad para su determinación, comparado con la clasificación 3+.

4.2.2 Análisis de variabilidad para RE y RP

Como resultado de la evaluación realizada por los patólogos a través de los formularios on-line, de las 10 placas virtuales evaluadas para cada receptor hormonal (RE y RP), se obtuvo un total de 90 ROIs con su respectivo diagnóstico y 30 evaluaciones generales, las que fueron expresadas en porcentaje de positividad nuclear para cada marcador respectivo. De manera similar a lo observado para HER2, los resultados mostraron la existencia de variabilidad en: i) tamaño de las ROIs, ii) localización de ROIs y iii) evaluación del estado de positividad, tanto para RE como para RP.

Del proceso de identificación y etiquetado de núcleos se obtuvieron 15 ROIs evaluadas, las que fueron utilizadas para configurar el algoritmo de cuantificación nuclear, tanto para RE como para RP. A continuación, se describen en detalle los resultados del análisis de ROIs para ambos marcadores.

4.2.2.1 Variabilidad inter-observador en la identificación nuclear

El proceso de etiquetado de núcleos permitió evidenciar que el número de núcleos positivos fluctuó entre valores cercanos, es decir, hubo baja variabilidad, mientras que para los núcleos negativos la variabilidad observada fue mayor (ver Anexo VI). La causa de esta variabilidad puede ser atribuida a una tinción tenue e inespecífica presentada por los núcleos, similar a la tinción presentada por el estroma o tejido de fondo, lo que dificulta la detección de núcleos negativos, exigiendo a los patólogos un mayor esfuerzo y la aplicación de criterios de interpretación particulares al momento de realizar el etiquetado.

4.2.2.2 Heterogeneidad en receptores hormonales

La variabilidad en el etiquetado de núcleos descrita en el punto anterior sugiere la existencia de heterogeneidad en la tinción en la muestra, factor que incidiría en cierto grado a la evaluación realizada por el patólogo que pone de manifiesto criterios propios de interpretación. En referencia a lo mencionado, para el marcador RP se observaron dos casos particulares: En el primero, un patólogo evaluó una placa virtual (RP_0005) con 95% de positividad, un segundo la evaluó con 0%. Más aún, un tercer patólogo describió este mismo caso como “placa no evaluable”. En el segundo caso (placa RP_0011), el porcentaje de positividad fluctuó entre 30 a 100%. Con el propósito de verificar el resultado, se solicitó a uno de los tres patólogos realizar una evaluación extra. Se reveló una distribución muy heterogénea de la marca dentro del tejido, lo que dificultó la evaluación. En la Sección 4.4 y 4.5 se visualizan los resultados de la evaluación realizada por los tres patólogos.

El *gold standard* generado permitió la configuración específica de los algoritmos predefinidos del *software* NDP.analyze, tanto para la cuantificación de HER2 como para los receptores hormonales, a través de pruebas iterativas de ajuste con el fin de adaptarlos a la realidad local. Los resultados obtenidos permitieron establecer parámetros tanto para el algoritmo conectividad de membrana como para el nuevo algoritmo de cuantificación nuclear RG, como se describió en la Sección 3.3. Así

mismo, se siguieron metodologías para la traducción de los valores cuantitativos a la forma tradicional de reporte ASCO/CAP. Una vez configurados, los algoritmos se utilizaron para la segmentación y cuantificación de las ROIs provistas por los patólogos.

4.3 CUANTIFICACIÓN IHQ EN HER2 Y ANÁLISIS ESTADÍSTICO DE CONCORDANCIA

La conversión de los valores de conectividad (escala continua de 0 a 1) a la clasificación ASCO/CAP (0/1+, 2+ y 3+) fue realizada a través de valores de corte (*cut-off*), determinados a través del entrenamiento de clasificadores de ML (ver Sección 3.3.1.1). Posteriormente estos clasificadores ya entrenados fueron utilizados en el conjunto de validación. Los resultados de los métodos empleados se describen a continuación.

4.3.1 Conversión de valores de conectividad mediante el clasificador CART

Los valores de corte fueron determinados a partir de 108 ROIs del conjunto de entrenamiento. Como se mencionó en el Capítulo 3.3.1.1, este clasificador fue entrenado con valores de conectividad obtenidos a 75% de sensibilidad, siguiendo la metodología propuesta por los desarrolladores del algoritmo (Visiopharm). Los resultados del entrenamiento de este clasificador (ver Anexo VII) tuvieron una precisión del 92%, con 99 ROIs clasificadas correctamente y una concordancia $k = 0.87$ respecto al diagnóstico de los patólogos, que corresponde a un nivel casi perfecto. Los valores de corte obtenidos para la conversión de valores de conectividad a la clasificación ASCO/CAP (ver Figura 4.7B), fueron aplicados a 117 ROIs del conjunto de validación (ver Anexo VIII). La precisión de la clasificación observada entre la clasificación entregada por los patólogos y la realizada por CART fue del 62 %, con una clasificación similar a la de los patólogos en 73 ROIs (clasificación correcta) y una concordancia $k = 0.38$. Los resultados del rendimiento del clasificador junto a la matriz de confusión de la clasificación correcta e incorrecta se ilustran en la Anexo VII.

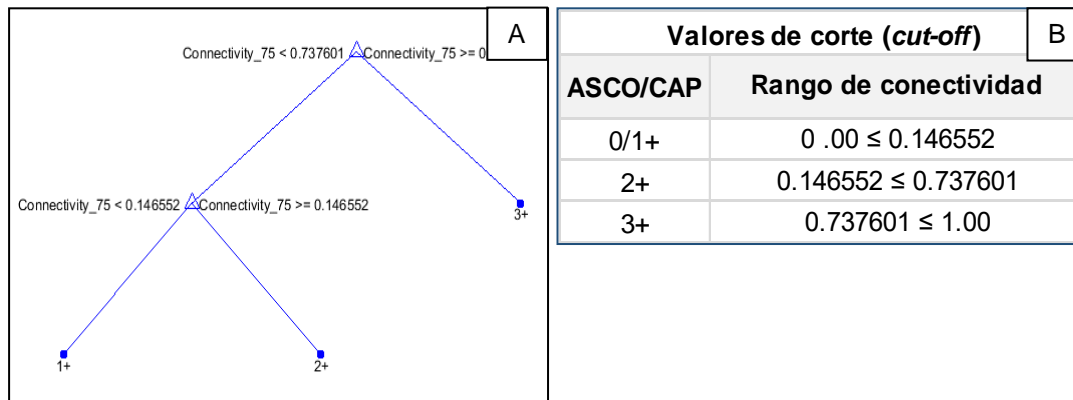


Figura 4.7: Traducción de valores de conectividad a clasificación ASCO/CAP mediante metodología CART. A) Representación gráfica del clasificador mediante árboles de decisión, cada nodo se encuentran los valores de corte para cada clasificación ASCO/CAP (0/1+, 2+ y 3+); B) Conversión de valores de conectividad en escala continua (0-1) a una escala discontinua para HER2 según clasificación ASCO/CAP.

A partir de las 117 ROIs clasificadas, se obtuvieron las clasificaciones generales por placa virtual, la que se presenta en la Tabla 4.1. Las clasificaciones 0+ y 1+, ambas negativas se consideraron como 0/1+. En las filas destacadas en color rojo se observa una diferencia entre la clasificación realizada por los patólogos y la realizada por el clasificador *CART* obtenidos a partir de los valores de conectividad de las ROIs. Se observa la existencia de discordancia en 6* de 14 placas en las cuales el clasificador discrepó con el voto de mayoría de los patólogos.

Tabla 4.1: Clasificación general realizada por patólogos versus CART.

Nº	Slide ID	Final_class_pat			Majority vote	Class_CART (0/1+,2+,3+)
		P1	P2	P3		
1	HER2_0001	2+	2+	2+	2+	2+
2	HER2_0002	0/1+	0/1+	0/1+	0/1+	0/1+
3	HER2_0006	2+	2+	0/1+	2+	2+
4*	HER2_0008	0/1+	2+	0/1+	0/1+	2+
5*	HER2_0009	0/1+	2+	2+	2+	0/1+
6	HER2_0010	0/1+	0/1+	0/1+	0/1+	0/1+
7*	HER2_0011	0/1+	2+	2+	2+	0/1+
8	HER2_0012	0/1+	0/1+	0/1+	0/1+	0/1+
9	HER2_0014	2+	2+	2+	2+	2+
10*	HER2_0017	0/1+	2+	2+	2+	0/1+
11	HER2_0019	3+	3+	3+	3+	3+
12*	HER2_0020	0/1+	2+	2+	2+	0/1+
13	HER2_0025	0/1+	2+	0/1+	0/1+	0/1+
14*	HER2_0027	0/1+	2+	0/1+	0/1+	2+

4.3.1.1 Concordancia patólogos versus CART

Los resultados del análisis de concordancia observada entre la clasificación mediante estimación visual realizada por los 3 patólogos versus la clasificación mediante CART en 117 ROIs y posterior clasificación general del conjunto de validación, se presentan en las Tablas 4.2-4.

En la Tabla 4.2 se observan las concordancias en la clasificación de ROIs por patólogo (P1, P2 y P3). 9 ROIs de P1 fueron excluidas, debido a que presentaban diagnóstico “no evaluable” (muestra considerada sub-óptima para su evaluación). Se observa además un nivel de concordancia moderado ($k= 0.43 - 0.52$) entre los patólogos P1 (33 ROIs) y P2 (42 ROIs) respecto a CART con nivel de significancia alto, no se observa diferencia significativa. No obstante, P3 presenta un nivel de concordancia leve ($k = 0.09$) y si se observa diferencia significativa.

Tabla 4.2: Concordancia observada en ROIs distribuidas por patólogo

Concordancia patólogos vs CART en clasificacion de ROIs				
N° de casos	% Concordancia	Kappa	Error típ. asint. ^a	p-value
P1 vs CART (33 ROIs)	72.73%	.517	0.116	.000
P2 vs CART (42 ROIs)	66.67%	.434	.114	.000
P3 vs CART (42 ROIs)	47.62%	.089	.147	.469
Nivel de significancia	● < 0.1	**	< 0.01	
	* < 0.05	***	< 0.005	

En la Tabla 4.3 se observa la concordancia existente en las clasificaciones generales, patólogos versus CART oscilan entre leve y considerable, con niveles de significancia altos respecto a P1 ($k = 0.720$) y P2 ($k = 0.421$), no obstante, entre P1 y CART se observa diferencia significativa ($k = 0.1$). La concordancia observada entre el voto de mayoría de los patólogos y CART presenta un $k = 0.250$, con una diferencia significativa leve.

Tabla 4.3: Concordancia observada patólogos versus CART en clasificaciones generales

Concordancia patólogos vs CART en clasificaciones generales				
N° de casos	% Concordancia	Kappa	Error típ. asint. ^a	p-value
P1 vs CART (14)	85.71%	.720	.180	.001
P2 vs CART (14)	64.29%	.421	.181	.015
P3 vs CART (14)	50.00%	.101	.272	.650
Voto de mayoria vs CART (14)	57.14%	.250	.251	.247
Nivel de significancia	● < 0.1	**	< 0.01	
	* < 0.05	***	< 0.005	

Dado que mediante el uso del clasificador CART, que sigue la metodología propuesta por el desarrollador del algoritmo (Visiopharm), se obtuvo un valor relativamente bajo de concordancia respecto al diagnóstico entregado por los patólogos, se procedió al uso del clasificador SMO.

4.3.2 Cuantificación de HER2 y traducción de los valores de conectividad mediante el clasificador SMO

A diferencia del método anterior, este clasificador fue entrenado con valores de conectividad de 108 Ros del conjunto de entrenamiento, obtenidos en el rango de 60-90% de sensibilidad, mencionado en la Sección 3.3.1.1. El resultado del entrenamiento del modelo SMO (ver Anexo IX) obtuvo un 93.52% de exactitud, con 101 ROIs clasificadas correctamente y una concordancia $k = 0.895$ respecto al diagnóstico de los patólogos, lo que corresponde a un nivel casi perfecto indicando que se obtuvo una buena separación entre las clases (ver Figura 4.8A). El clasificador SMO ya entrenado clasificó las 117 ROIs del conjunto de validación (ver Anexo IX). La exactitud de la clasificación observada entre la clasificación entregada por los patólogos y la realizada por SMO (Figura 4.8B) fue del 69,23% (porcentaje de concordancia), con una clasificación similar a la de los patólogos en 73 ROIs (clasificación correcta). Los resultados del número de ROIs clasificadas correcta e incorrectamente se ilustran en la Figura 4.8.

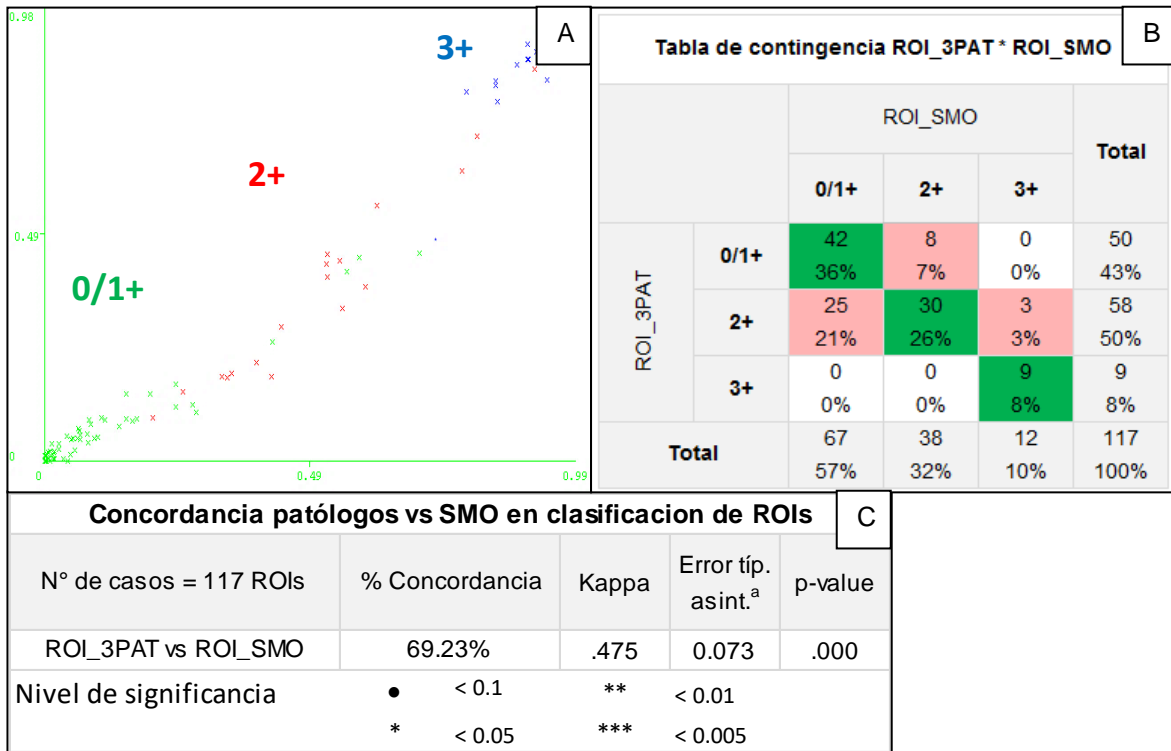


Figura 4.8: Traducción de valores de conectividad a clasificación ASCO/CAP mediante metodología SMO. A) Representación gráfica del clasificador SMO separando clases (0/1+, 2+ y 3+), proceso de conversión de valores conectividad de una escala continua (0-1) a una escala discontinua mediante separación de clases; B) Tabla de contingencia entre la evaluación visual realizada por 3 patólogos (117 ROIs del conjunto de validación) versus la clasificación realizada por el clasificador SMO. Se observa en cuadros verdes el número de ROIs concordantes y en cuadros rosados el número de ROIs discordantes. Se observa además que ROIs clasificadas como 0/1+ (25 ROIs) y 2+ (8 ROIs) presentan mayor discrepancia respecto a la clasificación de los patólogos; C) Resultados estadísticos del rendimiento del clasificador CART. Se observa una exactitud de 62.39% (porcentaje de concordancia), con valor de concordancia $k = 0.475$, nivel moderado y significancia alto, no se observa diferencia significativa.

A partir de las 117 ROIs clasificadas, se obtuvo una clasificación general por caso, la que se presenta en la Tabla 4.4, junto al voto de mayoría de los patólogos (es decir, el valor más frecuente observado) y el resultado del clasificador. La clasificación ASCO/CAP 0+ y 1+, ambas negativas se consideraron como 0/1+, En las filas destacadas en color rojo se observa una diferencia entre la clasificación realizada por los patólogos y la realizada por el clasificador SMO obtenidos a partir de los valores de conectividad de las ROIs. Se observa la existencia de discordancia en 4* de 14 placas en las cuales el clasificador discrepó con el voto de mayoría de los patólogos.

Tabla 4.4: Comparación de resultados de la clasificación general realizada por patólogos versus SMO.

Nº	Slide ID	Final_class_pat			Majority vote	Class_SMO (0/1+,2+,3+)
		P1	P2	P3		
1	HER2_0001	2+	2+	2+	2+	2+
2	HER2_0002	0/1+	0/1+	0/1+	0/1+	0/1+
3	HER2_0006	2+	2+	0/1+	2+	2+
4	HER2_0008	0/1+	2+	0/1+	0/1+	0/1+
5*	HER2_0009	0/1+	2+	2+	2+	0/1+
6	HER2_0010	0/1+	0/1+	0/1+	0/1+	0/1+
7*	HER2_0011	0/1+	2+	2+	2+	0/1+
8	HER2_0012	0/1+	0/1+	0/1+	0/1+	0/1+
9	HER2_0014	2+	2+	2+	2+	2+
10*	HER2_0017	0/1+	2+	2+	2+	0/1+
11	HER2_0019	3+	3+	3+	3+	3+
12*	HER2_0020	0/1+	2+	2+	2+	0/1+
13	HER2_0025	0/1+	2+	0/1+	0/1+	0/1+
14	HER2_0027	0/1+	2+	0/1+	0/1+	0/1+

Se observa discordancia en el voto de mayoría (P2 y P3) en casos con clasificación 2+ para los cuales el clasificador SMO dio una clasificación 0/1+ concordando con el tercer patólogo (P1).

4.3.2.1 Concordancia patólogos versus SMO

Los resultados del análisis de concordancia observada entre la clasificación mediante estimación visual realizada por los 3 patólogos versus la clasificación mediante SMO en 117 ROIs y posterior clasificación general del conjunto de validación, se presentan en las Tablas 4.5 y Tabla 4.7. En la Tabla 4.6 se presentan las concordancias observadas entre patólogos.

Los resultados del clasificador SMO fueron contrastados con la clasificación realizada por los patólogos tanto en ROIs como en las clasificaciones generales. En la Tabla 4.5 se muestran las concordancias observadas en ROIs del conjunto de validación (P1= 33 ROIs, P2 = 42 ROIs y P3 = 42), patólogos versus clasificador SMO, cuyos valores corresponden a un nivel entre aceptable y considerable. Una mayor concordancia es observada entre el clasificador SMO respecto a P1 ($k = 0.67$, 81.82%) y P2 ($k = 0.46$, 69.1%), no obstante, en P3 ($k = 0.28$, 59.52%) presenta una concordancia más baja, sin embargo, no se observa diferencia significativa (p-value < 0.05).

Tabla 4.5: Concordancia observada en ROIs distribuidas por patólogo

Concordancia patólogos vs SMO en clasificación de ROIs				
N° de casos	% Concordancia	Kappa	Error típ. asint. ^a	p-value
P1 vs SMO (33 ROIs)	81.82%	.667	0.119	.000
P2 vs SMO (42 ROIs)	69.05%	.457	.118	.000
P3 vs SMO (42 ROIs)	59.52%	.277	.139	.021
Nivel de significancia	● < 0.1	**	< 0.01	
	* < 0.05	***	< 0.005	

En la Tabla 4.6, se presentan las concordancias inter-observador en la clasificación general del conjunto de validación (14 placas). Los niveles de concordancia se encuentran entre aceptable y moderado, con niveles de significancia considerados altos. Destaca una mayor concordancia entre P2 y P3 ($k = 0.51$), sin embargo, se observa una concordancia mucho menor entre P1 respecto a P3 ($k = 0.27$).

Tabla 4.6: Resultados de la prueba de concordancia *Cohen's Kappa* entre tres patólogos (P1-3).

Concordancia inter-observador en clasificaciones generales				
N° de casos	% Concordancia	Kappa	Error típ. asint. ^a	p-value
P1 vs P2 (14)	50.00%	.274	0.150	.044
P1 vs P3 (14)	64.29%	.346	.232	.084
P2 vs P3 (14)	71.43%	.509	.190	.007
Nivel de significancia	● < 0.1	**	< 0.01	
	* < 0.05	***	< 0.005	

Los resultados de la concordancia patólogos versus el clasificador SMO en la clasificación de 14 placas, se presentan en la Tabla 4.7. Para P1 se obtuvo un nivel casi perfecto ($k = 1$), respecto a P2 y P3 niveles aceptables ($k = 0.276$ y 0.346 , respectivamente). La concordancia observada entre el voto de mayoría de los patólogos y SMO presenta un $k = 0.509$, no se observa diferencia significativa ($p < 0.01$).

Tabla 4.7: Concordancia patólogos versus SMO

Concordancia patólogos vs SMO en clasificaciones generales				
N° de casos	% Concordancia	Kappa	Error típ. asint. ^a	p-value
P1 vs SMO (14)	100.00%	1.000	.000	.000
P2 vs SMO (14)	50.00%	.274	.150	.044
P3 vs SMO (14)	64.29%	.346	.232	.084
Voto de mayoría vs SMO (14)	71.43%	.509	.190	.007
Nivel de significancia	● < 0.1	**	< 0.01	
	* < 0.05	***	< 0.005	

La mayor discordancia observada fue en casos clasificados como 2+ tanto en la clasificación entregada por los patólogos como en la clasificación mediante los dos clasificadores utilizados (CART y SMO). Sin embargo, los valores de las concordancias observadas en la clasificación tanto por ROI como por clasificación general por placa indican que el método propuesto mediante el uso del algoritmo conectividad de membrana junto a clasificadores de máquinas de aprendizaje resultan equivalentes al diagnóstico realizado por los patólogos.

Finalmente, resultados del segundo método (SMO) mostraron valores de concordancia mayores a los observados con el método CART, llegando a nivel moderado ($k = 0.47$), comparando las 117 ROIs del conjunto de validación. En la clasificación final las concordancias observadas entre el clasificador respecto a cada patólogo obtuvieron niveles de concordancia aceptable y casi perfecta (P1-SMO, $k = 1$; P2-SMO, $k = 0.274$; P3-SMO, $k = 0.346$). Así mismo, la concordancia observada

entre el voto de mayoría de los patólogos versus el clasificador tuvo un nivel moderado ($k = 0.51$), con 10 de 14 placas virtuales concordantes y 4 discordantes. Por otro lado, como se mencionó anteriormente la concordancia observada entre los patólogos (P1-P2, $k = 0.27$; P1-P3, $k = 0.34$; P2-P3, $k = 0.51$), presentan niveles aceptables y moderados, en algunos casos los resultados se presentan relativamente más bajos a los observados entre el clasificador SMO.

4.4 CUANTIFICACIÓN IHQ EN RE Y ANÁLISIS ESTADÍSTICO DE CONFIABILIDAD

Para la obtención del porcentaje de positividad mediante el algoritmo “RG”, fue necesario el análisis estadístico de normalización de los coeficientes de positividad obtenidos de las muestras control positivo, previo a su aplicación y uso en el análisis de ROIs entregados por los patólogos, a fin de dar robustez a la nueva metodología propuesta. A continuación, se presentan los resultados de estos análisis, la cuantificación realizada y su posterior análisis estadístico de concordancia, tanto para RE como para RP.

4.4.1 Análisis coeficiente de positividad de muestras control para RE

Los coeficientes de positividad obtenidos del análisis de las 10 ROIs en las muestras control positivo para cada placa de RE, presentan una distribución normal (ver Tabla 4.8, Shapiro-Wilk). Los promedios de estos valores (cp100) también siguen una distribución normal (Tabla 2, línea amarilla inferior). Del mismo modo, los valores del conjunto de ROIs de todas las placas también siguen una distribución normal, una vez eliminadas las placas anormales RE_0001 y RE_002 (Tabla 2, última línea y Figura 4.12 C). Estas placas resultaron ser estadísticamente diferentes en relación a los demás valores según el test paramétrico Tukey's t-test ($p < 2,6E-15$) y el test no paramétrico Kolmogorov-Smirnov ($p < 3,2E-1$).

Tabla 4.8: Coeficientes RE de 100% positividad de 10 ROIs (cc_1-10) y promedios (cp₁₀₀) en 10 placas

Slide ID	cc_1	cc_2	cc_3	cc_4	cc_5	cc_6	cc_7	cc_8	cc_9	cc_10	cp ₁₀₀	p-val (SW)
RE_0001*	0,59	0,63	0,61	0,59	0,66	0,70	0,68	0,63	0,66	0,59	0,63	0,37
RE_0002*	0,57	0,54	0,53	0,55	0,55	0,54	0,55	0,59	0,57	0,52	0,55	0,61
RE_0003	0,40	0,41	0,40	0,37	0,40	0,35	0,38	0,38	0,33	0,35	0,38	0,58
RE_0004	0,33	0,42	0,35	0,38	0,38	0,43	0,43	0,39	0,38	0,38	0,39	0,45
RE_0005	-	0,30	0,32	0,31	-	-	0,32	0,33	0,30	0,29	0,31	0,71
RE_0006	0,48	0,44	0,51	0,38	0,40	0,49	0,33	0,40	0,49	0,57	0,45	0,96
RE_0008	-	-	0,54	0,43	0,44	0,43	0,51	0,48	0,52	0,51	0,48	0,20
RE_0009	0,45	0,43	0,47	0,35	0,42	0,42	0,38	0,43	0,43	-	0,42	0,36
RE_0010	0,48	0,44	0,45	0,43	0,39	0,42	0,45	0,41	0,40	0,43	0,43	0,95
RE_0011	0,49	0,45	0,49	0,49	0,46	0,42	0,35	0,39	0,42	0,36	0,43	0,22
p-val (SW)	0,0014 (*0,34)										0,61	

La columna p-val (SW) muestra los resultados del test de normalidad de Shapiro-Wilk (SW). Placas marcadas con (*) no fueron considerados por los resultados obtenidos con el *Tukey's Honest Significance Difference test (Tukey's HSD test)*. La última línea muestra los resultados p-val (SW) para todos los coeficientes RE de 100% positividad con (*) y sin RE_0001 y 2 juntos a los p-val (SW) de los promedios (cp₁₀₀). Aplicando un nivel de significancia de $\alpha = 0,1$, todos los valores pertenecen a una distribución normal.

La revisión visual de dichas placas realizada por un patólogo confirma lo indicado por la estadística. Se apreció una tinción más intensa que en las restantes, lo que conlleva a una mayor segmentación de núcleos positivos y valores mayores de cp₁₀₀. Excluyendo las placas RE_0001/2, el promedio de los coeficientes RE es 0,41 con un error estándar 0,01 y una desviación estándar 0,06.

Dicho promedio concuerda con el centro de una curva de Gauss ($y = I * \exp(-(x-\mu)^2/2\sigma^2)$) que ajustamos después de eliminar las placas con problemas (Figura 4.9 B/C). Se concluye que es posible definir un valor de referencia cp₁₀₀ alrededor de 0,41 que represente 100% de positividad con un error de 0,01 basado en las 8 placas restantes.

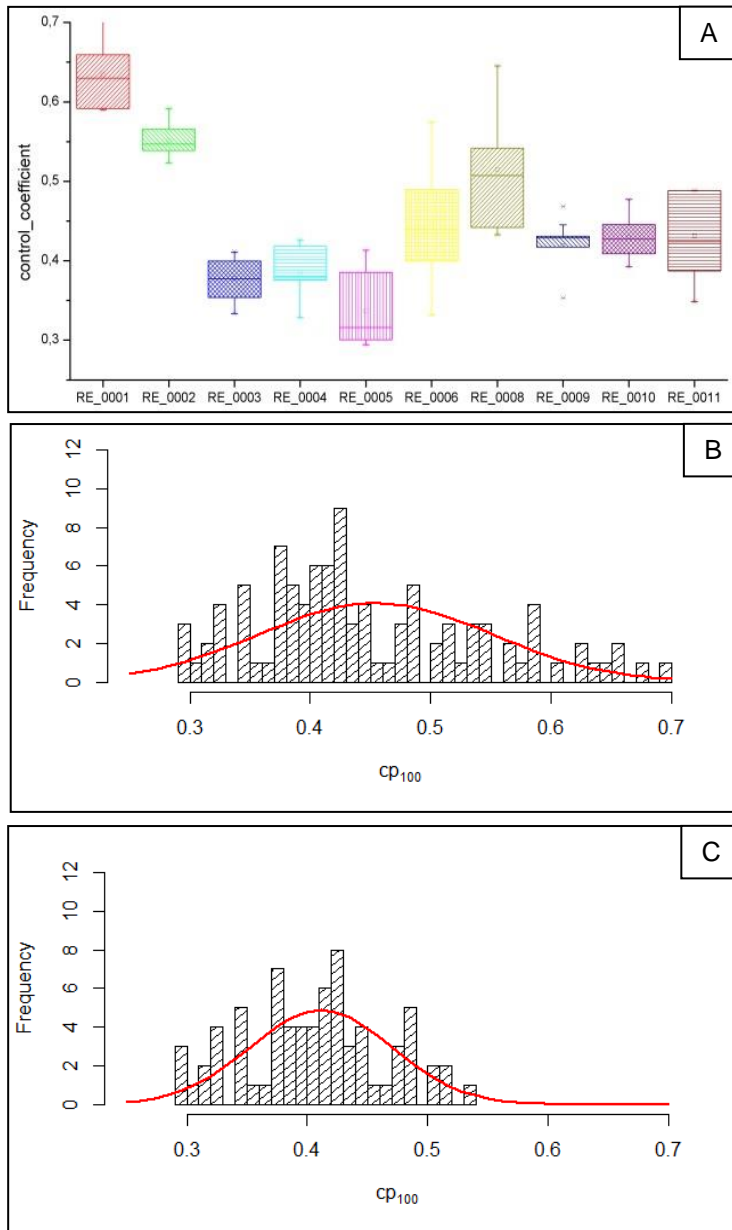


Figura 4.9: Variabilidad de coeficientes RE entre ROIs y placas (datos de Tabla 4.8). A) Barras muestran promedios de coeficientes RE, valores anormales (*outliers*) fueron marcados con (x) y se definen como 3/2 de la distancia entre los límites superiores e inferiores de la caja (box) y los promedios. Las barras de errores marcan los valores máximo y mínimo de todos los valores, sin considerar valores anormales. La caja define los cuartiles 25% superior e inferior de los datos, el cuadrado define los promedios y la línea del centro el mediano. B) Histograma de los coeficientes RE y curva de Gauss: $y = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$ con $\mu = 0,45$ y $\sigma = 0,09$. C) Histograma de los coeficientes RE excluyendo RE_0001/2 y curva de Gauss con $\mu = 0,41$ y $\sigma = 0,06$.

4.4.2 Porcentaje de positividad nuclear en RE

De las 90 ROIs obtenidas a partir de la evaluación de IHQ nuclear para RE en la etapa de generación del *gold standard*, se excluyeron 2 ROIs que fueron catalogadas como no evaluables por los patólogos. Estas fueron analizadas con el algoritmo RG para obtener su respectivo porcentaje de positividad nuclear (n_{pos}). Para dicho cálculo se utilizó el promedio del coeficiente de positividad de la muestra control, de

acuerdo a la Ecuación 3.1 (ver Capítulo 3 Sección 3.3.2.2. El análisis de cada ROI (ver Anexo X) permitió obtener el porcentaje de positividad general por placa, la que se presenta en la Figura 4.9.

Tabla 4.9: Comparación de la evaluación del porcentaje de núcleos positivos para RE realizada por tres patólogos (P1-3) versus el algoritmo GR (np_{pos}).

Placa	Patólogos P1-3 [%]			Promedio P1-3 [%]	RG [%] (np_{pos})
	P1	P2	P3		
RE_0001	90	100	100	97	83
RE_0002	100	100	80	93	100
RE_0003	100	100	100	100	100
RE_0004	100	90	100	97	100
RE_0005	100	100	100	100	100
RE_0006	100	100	70	90	99
RE_0008	95	93	90	93	93
RE_0009	0	0	0	0	0
RE_0010	100	93	100	98	94
RE_0011	100	90	90	93	74

4.4.3 Análisis de confiabilidad inter-patólogos y patólogos versus el algoritmo RG en RE

La Tabla 4.10 muestra los resultados de la prueba de concordancia K-alpha inter-observadores (tres patólogos) en base a las 10 placas. Los resultados muestran que existe una alta confiabilidad entre los tres expertos.

Tabla 4.10: Resultados del test de concordancia K-alpha entre tres patólogos (P1-3).

Patólogos P1-3	P1 vs P2	P2 vs P3	P3 vs P1
Krippendorff's K-alpha	0,994	0,975	0,974
LI 95%	0,991	0,943	0,941
LS 95%	0,998	0,998	0,997

LI: Límite inferior de confianza; LS: Límite superior de confianza. Análisis con 10 placas.

Se realizó además un test de concordancia K-alpha entre los tres patólogos y el algoritmo RG en base a las mismas 10 placas analizadas (Tabla 4.11). Los resultados indican que existe una alta confiabilidad entre los patólogos y el algoritmo RG.

Tabla 4.11: Resultados de concordancia K-alpha de tres patólogos (P1-3) versus el resultado del algoritmo

Patólogos vs Algoritmo RG	P1 vs RG	P2 vs RG	P3 vs RG	P1-3 vs RG
Krippendorff's K-alpha	0,980	0,983	0,962	0,994
LI 95%	0,947	0,964	0,930	0,987
LS 95%	0,999	0,999	0,990	0,999

RG.

LI: Límite inferior de confianza; LS: Límite superior de confianza. Análisis con 10 placas.

Finalmente se comprobó la concordancia K-alpha para las 88 ROIs evaluadas por los patólogos y la evaluación mediante RG (Tabla 4.12). Los resultados indican que existe una alta concordancia entre los patólogos y el algoritmo RG.

Tabla 4.12: Resultados de concordancia K-alpha de tres patólogos (P1-3) versus el resultado del algoritmo

Patólogos vs Algoritmo RG en 88 ROIs	P1 vs RG n=28	P2 vs RG n=30	P3 vs RG n=30	P-3 vs RG n=88
Krippendorff's K-alpha	0,981	0,977	0,966	0,974
LI 95%	0,966	0,957	0,944	0,962
LS 95%	0,994	0,992	0,984	0,984

LI: Límite inferior de confianza; LS: Límite superior de confianza.

En conclusión, los análisis estadísticos realizados muestran que hay una alta confiabilidad entre las evaluaciones realizadas por los patólogos participantes, por una parte, y por otra también muestran la existencia de una alta confiabilidad al comparar las evaluaciones realizadas por los patólogos y las efectuadas por nuestro algoritmo RG.

4.1 CUANTIFICACIÓN IHQ EN RP Y ANÁLISIS ESTADÍSTICO DE CONFIABILIDAD

Para la obtención del porcentaje de positividad mediante el algoritmo RG, fue necesario realizar análisis estadísticos de normalización de los coeficientes de positividad obtenidos de las muestras control positivo. Esto fue realizado en forma previa a su aplicación y uso en el análisis de las ROIs entregadas por los patólogos, con el fin de dar robustez a la nueva metodología propuesta (ver Sección 3.3.5).

A continuación, se presentan los resultados de estos análisis, la cuantificación realizada y su posterior análisis estadístico de concordancia, tanto para RE como para RP.

4.1.1 Análisis del coeficiente de positividad de muestras control para RP

De manera similar a los resultados obtenidos para los coeficientes RE, los coeficientes RP de las 10 ROIs de control positivo presentan distribuciones normales (Tabla 4.13, columna amarilla Shapiro-Wilk). Los valores promedios de las 10 ROIs dentro de cada placa también presentan una distribución normal (ver Tabla 4.13, línea inferior amarilla). El conjunto de valores de todos los ROIs también muestran una distribución normal, una vez que se excluye la placa anormal RP_0006 (Tabla 4.13, * última línea y Figura 4.10 B/C). Los valores de dicha placa fueron evaluados estadísticamente como diferente de los demás según el test paramétrico Tukey's *t*-test ($p < 0,005$) y el test no-paramétrico Kolmogorov-Smirnov ($p < 0,004$).

Tabla 4.13: Coeficientes RP de 100% positividad de 10 ROIs (cc_1-10) y promedios (cp₁₀₀).

Slide ID	cc_1	cc_2	cc_3	cc_4	cc_5	cc_6	cc_7	cc_8	cc_9	cc_10	cp ₁₀₀	p-val (SW)
RP_0001	0,42	0,44	0,46	0,49	0,41	0,44	0,44	0,44	0,45	0,44	0,44	0,31
RP_0002	0,47	0,49	0,48	0,50	0,51	0,49	0,49	0,45	0,43	0,46	0,48	0,35
RP_0003	0,43	0,42	-	0,44	0,43	0,46	0,42	0,46	0,44	0,44	0,44	0,59
RP_0004	0,50	0,47	0,45	0,48	0,46	0,45	0,48	0,47	0,46	0,43	0,46	0,98
RP_0005	-	0,43	0,42	0,49	0,45	0,43	-	0,49	0,52	0,49	0,47	0,22
RP_0006*	-	0,48	0,55	0,55	0,46	0,52	0,56	0,46	-	0,53	0,51	0,57
RP_0007	0,42	0,45	0,47	0,45	0,45	0,48	0,42	0,45	0,44	0,44	0,45	0,86
RP_0008	0,55	0,53	0,52	0,51	0,49	0,52	0,51	0,47	0,49	0,49	0,51	0,86
RP_0009	0,48	-	0,47	0,48	0,45	0,46	0,47	0,47	0,44	0,47	0,47	0,71
RP_0011	0,44	0,47	0,52	0,45	0,51	0,55	0,51	0,50	0,48	0,42	0,49	0,87
p-val (SW)	0,002 (*0,17)										0,58	

La columna p-val (SW) muestra los resultados del test de normalidad de Shapiro-Wilk. Las placas marcadas con (*) no fueron consideradas, de acuerdo a los resultados obtenidos con el *Tukey's Honest Significance Difference test (Tukey's HSD test)*. La última línea muestra los resultados p-val (SW) para todos los coeficientes RP de 100% positividad con (*) y sin RP_0006 juntos a los p-val (SW) de los promedios (cp₁₀₀). Aplicando un nivel de significancia de $\alpha = 0,1$, todos los valores pertenecen a una distribución normal.

Una revisión más acuciosa de la placa RP_0006 reveló una tinción levemente más intensa que la de las otras placas, lo que probablemente incidió en la mayor segmentación de núcleos positivos y mayores coeficientes cp₁₀₀ observados. Al eliminar esta placa, el promedio de todos los coeficientes RP es 0,47 con un error estándar de 0,01 y una desviación estándar de 0,03. Dicho valor promedio concuerda con el centro de la curva de Gauss ($y = I * \exp(-x - \mu)^2 / 2\sigma^2$) utilizada en el ajuste (Figura 4.10 B/C). Este análisis permite concluir que, basado en las 9 placas utilizadas, es posible definir un valor de referencia cp₁₀₀ alrededor de 0,47 que represente el 100% de positividad, con un error de 0,01.

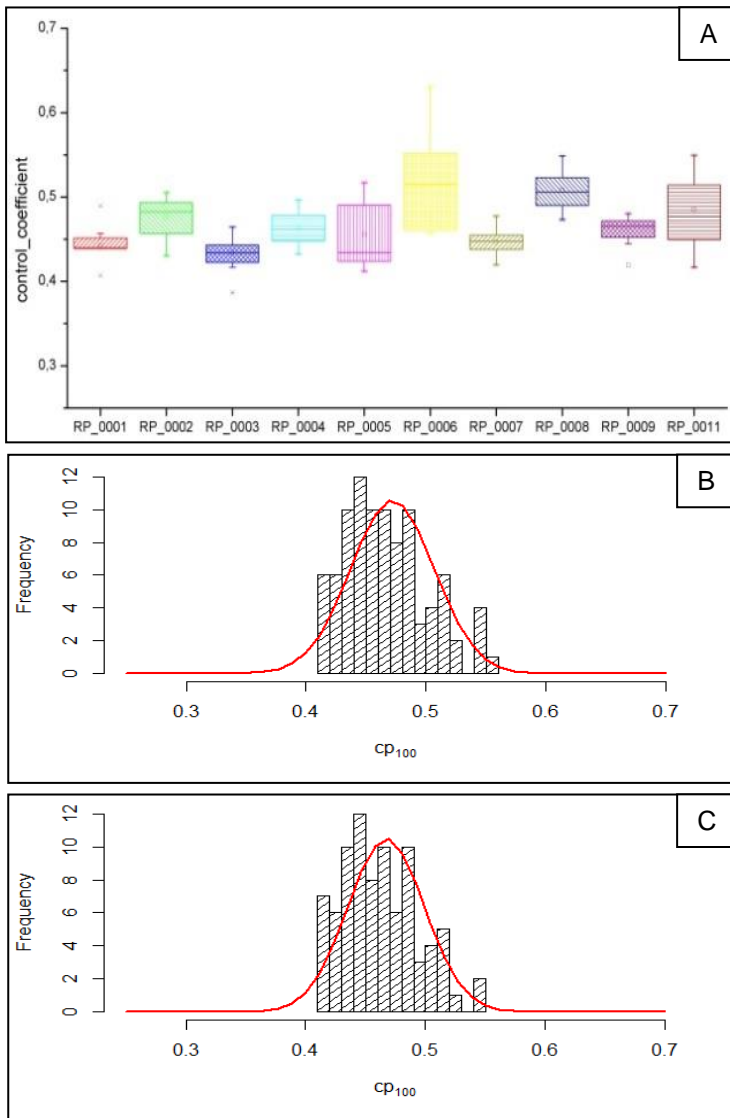


Figura 4.10: Variabilidad de coeficientes RP entre ROIs y placas (datos de Tabla 4.13). A) Barras muestran promedios de coeficientes RP, valores anormales (*outliers*) fueron marcados con (x) y se definen como $3/2$ de la distancia entre los límites superiores e inferiores de la caja (box) y los promedios. Las barras de errores marcan valores máximo y mínimo de todos los valores, sin considerar valores anormales. La caja define los cuartiles 25% superior e inferior de los datos, el cuadrado inserto en su interior define los promedios y la línea del centro corresponde a la mediana. B) Histograma de los coeficientes RP y curva de Gauss: $y = I * \exp(-(x - \mu)^2 / 2\sigma^2)$ con $\mu = 0.47$ y $\sigma = 0.04$. C) Histograma de los coeficientes RP excluyendo RP_0006 y curva de Gauss con $\mu = 0.46$ y $\sigma = 0.03$.

En aquellos casos en que la densidad de núcleos en la muestra bajo estudio fue mayor que la observada en la muestra control, considerados como 100%.

4.4.4 Porcentaje de positividad nuclear en RP

A partir de las 90 ROIs provistas por los patólogos en las 10 placas virtuales evaluadas, 8 fueron excluidas por ser consideradas no evaluables, quedando así un

total de 82 ROIs para el análisis y obtención de su respectivo porcentaje de positividad con el algoritmo RG (ver Anexo XI). De forma similar a lo realizado para RE, se obtuvo un porcentaje de positividad general por placa que se presenta en la Tabla 4.14.

Tabla 4.14: Tabla de evaluación de tres patólogos (P1-3) y el resultado del algoritmo RG para RP.

Placa	Tres patólogos P1-3 [%]			Promedio P1-3 [%]	RG [%] ($n_{p_{pos}}$)
	P1	P2	P3		
RP_0001	0	0	0	0	46
RP_0002	100	100	90	97	100
RP_0003	95	90	95	93	100
RP_0004	100	100	90	93	100
RP_0005*	-	95	0	32	83
RP_0006	90	100	100	97	100
RP_0007	0	0	0	0	6
RP_0008	0	0	0	0	10
RP_0009	40	55	60	52	90
RP_0011*	100	85	30	72	91

Los resultados demuestran que existe una importante variabilidad interobservador en la clasificación de las placas 5 y 11 (marcadas con asterisco). Una segunda revisión de ambas placas, realizada por un patólogo experto, reveló como causa de dicha variabilidad la existencia de una distribución muy heterogénea de la marca dentro del tejido. En particular, en la placa 5, dos patólogos dieron más importancia a zonas con poca marca (evaluados con 0%), mientras que el tercer patólogo dio más importancia a las zonas más marcadas (evaluadas con 95%). El algoritmo RG entregó un valor de 83% para la placa, apoyando la evaluación realizada por tercer patólogo. Una explicación similar se obtuvo para placa 11, en la cual los patrones de marcación fueron muy heterogéneos, resultando en importantes diferencias inter-observadores en la estimación de los porcentajes. El algoritmo RG

entregó un valor más bien alto (91%), apoyando los valores más altos (85 y 100%) asignados por dos de los tres patólogos, en contraste con la estimación más baja (30%) asignada por uno de ellos.

El comportamiento del algoritmo RG frente a situaciones donde existen placas con marcación heterogénea enfatiza la propiedad del algoritmo de detectar sectores con alta marcación y evaluar el porcentaje de marcación en dichas regiones, en preferencia a regiones o sectores con baja marcación.

4.4.5 Análisis de confiabilidad inter-patólogos y patólogos versus el algoritmo RG en RP

Los resultados del análisis estadístico K-alpha de la evaluación de RP por los patólogos se presenta en la Tabla 4.15. Los valores obtenidos excluyendo las placas RP_0005 y RP_0011 se marcan con (*). En general, los resultados demuestran que existe una confiabilidad alta, a pesar de los valores heterogéneos observados en las placas 5 y 11.

Tabla 4.15: Resultados de concordancia K-alpha para RP entre patólogos de sin RP_0005/11 (*).

Patólogos P1-3	P1 vs P2	P2 vs P3	P3 vs P1
<i>Krippendorff's K-alpha</i>	0,988	0,949	0,904
LI 95%	0,974	0,851	0,729
LS 95%	0,998	0,999	0,997
<i>Krippendorff's K-alpha*</i>	0.984*	0.998*	0.975*

LI: Límite inferior de confianza; LS: Límite superior de confianza.

Los resultados de confiabilidad K-alpha de la evaluación por placa de patólogos versus el algoritmo RG se muestran en la Tabla 4.16. Valores sin RP_0005/11 se marcan con (*). Se puede observar que los resultados se encuentran en el rango de 'No confiable'.

Tabla 4.16: Resultados de concordancia K-alpha para RP entre tres patólogos versus el resultado del algoritmo RG de los 10 casos; sin RP_0005/11 (*).

<i>Patólogos vs Algoritmo RG</i>	<i>P1 vs RG</i>	<i>P2 vs RG</i>	<i>P3 vs RG</i>	<i>P1-3 vs RG</i>
<i>Krippendorff's K-alpha</i>	0,175	0,190	0,155	0,200
<i>LI 95%</i>	0,070	0,059	0,059	0,053
<i>LS 95%</i>	0,995	0,964	0,844	0,923
<i>Krippendorff's K-alpha*</i>	0.330*	0.286*	0.289*	0.308*

LI: Límite inferior de confianza; LS: Límite superior de confianza.

La Tabla 4.17 presenta los resultados de confiabilidad K-alpha por ROIs evaluadas por los patólogos versus la obtenida mediante el algoritmo RG. Se obtiene un nivel de confiabilidad considerable ($K\text{-alpha} > 0,7$). Sin embargo al excluir las ROIs de placas (*) RP_0001, RP_0005 y RP_0011, se obtiene un nivel confiable ($K\text{-alpha} > 0,8$).

El resultado de una baja concordancia era esperado debido a las diferencias observadas entre patólogos para las placas RP_0005 y RP_0011, sin embargo al realizar una comparación por ROI el nivel de confiabilidad aumenta. De manera similar a RE se sugiere en el futuro ampliar el tamaño de la muestra a un valor cercano a 30 o más placas.

Tabla 4.17: Resultados de concordancia K-alpha para RP de tres patólogos versus el resultado del algoritmo RG en 82 ROIs analizados; sin RP_0001/5/11 (*) N = 59 ROIs.

<i>Patólogos vs Algoritmo RG</i>	<i>P1 vs RG</i>	<i>P2 vs RG</i>	<i>P3 vs RG</i>	<i>P 1-3 vs RG</i>
<i>Krippendorff's K-alpha</i>	0,679	0,766	0,669	0,774
<i>LI 95%</i>	0,253	0,501	0,390	0,647
<i>LS 95%</i>	0,949	0,979	0,896	0,888
<i>Krippendorff's K-alpha*</i>	0.846*	0.914*	0.859*	0.879

LI: Límite inferior de confianza; LS: Límite superior de confianza.

Los resultados de la cuantificación realizada tanto para RE como para RP mediante el algoritmo RG, permitió determinar el porcentaje de positividad nuclear en base al tejido control (100% positivo). Los coeficientes de positividad calculados en muestras con 100% de positividad generaron valores con distribuciones normales, no solo en los diferentes ROIs dentro de una placa, sino también entre las placas

(Shapiro-Wilk). El análisis estadístico realizado en los datos de calibración sugiere que estos presentan distribuciones normales (Gaussianas con valores σ relativamente bajos), hecho que nos permite concluir que el algoritmo RG refleja la variación natural entre muestras. Esto es explicable por la presencia de factores tales como, contenido de grasa en el tejido mamario, variaciones en el grosor del corte, etc.

En general los coeficientes de positividad de las muestras en estudio tanto para RE como RP pertenecen a distribuciones normales (ver Figuras 4.9 y 4.10, RE: $\mu = 0.41$, σ o SD = 0.06, SE = 0.006; RP: $\mu = 0.46$, σ o SD = 0.03, SE = 0.003). Las fluctuaciones dentro de la distribución normal se pueden explicar por los factores antes mencionados, así como por la reproducibilidad de la tinción dentro de un margen de variabilidad aceptable en el laboratorio. Finalmente, la metodología empleada si permite la obtención de porcentajes de positividad basados en la comparación con la muestra control, para verificar el rendimiento del algoritmo se recomienda en el futuro ampliar el número de muestras.

5.DISCUSIÓN

La constitución de CPDAI como primer centro chileno de PD y Microscopía virtual (MV) abrió un nuevo paradigma en el flujo de trabajo necesario para pasar de una forma manual a una digital, no solo en el campo de la AP sino también a nivel académico y de investigación, en Chile y la región. Por lo anterior, dentro de su labor le ha correspondido a CPDAI determinar, entre otras tareas, estrategias adecuadas para la inserción e implementación de la MV y PD en el flujo de trabajo de patólogos, académicos e investigadores.

La utilización de dichas herramientas ha sido ampliamente desarrollada e investigada y su adopción ha experimentado un fuerte crecimiento en países norteamericanos [82] y europeos [24]. En Chile, no obstante, aún es considerada una tecnología a explorar, en la que múltiples factores, tanto humanos como tecnológicos y de gestión influirán de manera importante en su adopción.

El trabajo presentado no solo estableció un proceso y protocolo para el uso de las herramientas de PD y su aplicación en el apoyo al diagnóstico clínico de patologías como el CM, sino que además permitió contextualizarlo a la realidad local. Se combinaron metodologías propuestas y descritas previamente en la literatura con metodologías de investigación empíricas, cuya discusión es abordada en este capítulo.

5.1 FLUJO DE TRABAJO DEL PROCESO DE VALIDACIÓN: CONECTANDO PATÓLOGOS Y ORGANIZACIONES

La notación gráfica utilizada (BPMN) para la diagramación y representación de las tareas y/o actividades que componen el flujo de trabajo del proceso de validación desarrollado en esta investigación, demostró ser una alternativa adecuada ya que favorece la comprensión y estructuración del proceso de forma clara y precisa [71]. Permite, por otra parte, dar seguimiento y monitoreo a la información que fluye en cada una de las etapas, beneficios mencionados en la literatura y que fueron descritos ampliamente en el Capítulo 2, Sección 2.6.

La decisión de representar el flujo de trabajo en la notación BPMN, ayudó a definir y estandarizar los procedimientos y/o protocolos de trabajo fundamentales para proceder con la validación de PD. Permitió además representar las tareas y/o actividades del proceso de validación, plasmando la comunicación interdisciplinaria existente entre profesionales y la asignación de responsabilidades [70, 71]. Bajo el mismo contexto, se observa la gestión desde tres perspectivas diferentes: i) la perspectiva CPDAI, ii) del patólogo y iii) la organización con la que se coordinó el proceso. Por tanto, las actividades y/o tareas establecidas que componen el proceso diseñado quedan como un expediente y un insumo valioso, pues se ha mostrado que, siguiendo la metodología propuesta, es posible gestionar un proceso de validación de los algoritmos de cuantificación y las aplicaciones de PD.

Por otra parte, el flujo de trabajo implementado constituirá una guía o pauta para realizar futuras validaciones de otros marcadores inmunohistoquímicos, en un contexto local, dado que la notación utilizada es suficientemente flexible para aceptar cambios y/o modificaciones.

5.2 PLACAS VIRTUALES Y GENERACIÓN DE UN *GOLD STANDARD*

Los beneficios de las placas virtuales fueron claramente evidenciados en el proceso de generación de *gold standard*, permitiendo: i) el acceso remoto y simultáneo de los especialistas a toda la muestra en estudio, sin límites de lugar o tiempo a diferencia del método tradicional que si requiere el envío de la placa de vidrio; ii) selección de múltiples ROIs facilitando la aplicación de algoritmos de análisis de imagen, que tiempo atrás esta se veía limitada a la captura de únicos campos de análisis y no así a toda la muestra en estudio [1] y por tanto se requería mayor tiempo de interacción con los especialistas o la intervención de terceros.

5.3 CUANTIFICACIÓN AUTOMATIZADA Y VARIABILIDAD EN HER2

Nuestro estudio pone en evidencia que la cuantificación IHQ de HER2 en membrana celular mediante análisis de imagen tiene un rendimiento equivalente a la

evaluación realizada por los patólogos (*gold standard*). Además, se puso a prueba el rendimiento del algoritmo “conectividad de membrana” del software NDP.analyze de Visiopharm, algoritmo que considera la medición de la intensidad de tinción y la continuidad de la membrana, entregando un valor cuantitativo de la membrana marcada o medición “*chicken wire*” método demostrado más preciso en comparación al uso de sólo la medición de intensidad de tinción en HER2 [7].

Debido a que la licencia del software adquirido es de uso en investigación, fue necesaria la adición de un proceso de conversión de los valores de conectividad (rango de 0-1) a la nomenclatura ASCO/CAP a diferencia de los estudios en los cuales el algoritmo contaba con el clasificador incorporado y/o contaban con puntos de corte definidos [4, 81], pero no aplicables a nuestro estudio por tratarse de diferentes muestras, debido al cual el algoritmo debía ser estandarizado. Como se detalló (Sección 2.4.4), esto fue factible de realizar a través de clasificadores de ML, por un lado, utilizando clasificadores de árboles de decisión como el CART, siguiendo la metodología propuesta por el desarrollador [7] y por el otro algoritmo de *Support Vector Machine* como el SMO.

La implementación del método CART permitió obtener rangos de conectividad para cada clasificación (0/1+, 2+ y 3+) [7] obtenidos para un único valor de sensibilidad. Las concordancias observadas tanto en ROIs como en el voto de mayoría de los patólogos ($k = 0.38$, $k = 0.25$) obtuvieron niveles aceptables. Si bien, se obtuvieron valores de corte para la conversión de los valores de conectividad a la clasificación ASCO/CAP (mapeo), los resultados de concordancia se muestran relativamente bajos respecto a los patólogos.

Por tanto, se determinó el uso del segundo método de clasificación, el cual consideró mayores “*features*” o valores de conectividad para el análisis y entrenamiento del clasificador a diferencia del primer método que solo consideró valores únicos de conectividad por ROI, como fue descrito en el Capítulo 4.3.1 [7]. La implementación del segundo método (SMO) mostró valores de concordancia mayores a los observados con el método CART, tanto en ROIs como en clasificaciones generales (voto de mayoría) con $k = 0.47$ y $k = 0.51$ respectivamente.

En este contexto esta metodología permitiría seguir el comportamiento de ROIs puestas en estudio, como fue ilustrada en los gráficos de curvas ilustradas en la Sección 3.3.1. En este sentido se destaca las bondades del algoritmo al entregar un valor numérico de las características de la membrana celular analizadas en una ROI, permitiendo la transformación de una evaluación cualitativa “*chicken wire*” a una cuantitativa.

Las clasificaciones 0/1+ y 3+ fueron en las que se observó mayor concordancia, tanto en la evaluación visual como mediante los algoritmos, lo que es apoyado por la literatura, donde se indica que estas categorías son discriminadas consistentemente, tanto por los patólogos como por los métodos automatizados [4, 81]. Los casos donde se observó discordancia fueron aquellos clasificados como 2+, los que son considerados casos complejos de evaluar, corroborando lo descrito en la literatura [4, 81, 86]. Estos valores resultan esperables y se justifican debido a la variabilidad observada en las ROIs en términos de localización, tamaño y clasificación, aspectos que influyen en la variabilidad de la evaluación del marcador HER2. Este hecho explicaría la detección de algunas placas evaluadas por los patólogos como 2+ y que el algoritmo SMO, que fue el que tuvo mayor concordancia, las clasificara principalmente como 0/1+. Cabe recordar que ambos clasificadores fueron entrenados a partir de la clasificación entregada por los patólogos (*gold standard*), utilizando placas en las que se observó una concordancia total.

Estos umbrales de configuración y traducción de los valores de conectividad a la nomenclatura ASCO/CAP brindan una cuantificación automatizada con resultados equivalentes a la realizada por los patólogos y con una concordancia mayor a la observada entre patólogos. Este estudio permite concluir que las herramientas de análisis de imagen, junto a clasificadores de máquinas de aprendizaje permiten objetivar la cuantificación para este marcador de membrana.

En forma paralela, nuestro estudio permitió cuantificar la variabilidad entre patólogos para la evaluación de HER2, confirmando lo que ha sido descrito en la literatura [4, 11, 81]. En este mismo contexto en Chile se llevó a cabo el año 2012 un estudio sobre la variabilidad del análisis del marcador HER2, que concluyó que “la

variabilidad de la evaluación del estado de HER2 en laboratorios nacionales es similar a lo descrito en la literatura internacional, llegando al 19,7% cuando se compara con la técnica de IHQ y alcanza un 26,9% cuando se compara con la técnica FISH”. Sin embargo este estudio no identifica las fuentes de variabilidad de HER2 [20]. En la actualidad, considerando las nuevas herramientas tecnológicas disponibles, como son la digitalización de placas, el acceso a placas virtuales y el análisis de imágenes, es plausible proponer que estas herramientas de PD ayudarían en gran medida a detectar dichas fuentes de variabilidad o los factores que influyen en la evaluación, de manera mucho más específica y precisa.

Las fuentes de variabilidad podrían ser atribuidas en gran medida a las ROIs que seleccionan los patólogos al momento de evaluar una placa, lo que determinaría la presencia o no de heterogeneidad, tanto en la tinción como en los criterios de interpretación [4, 48, 81]. Dado que los casos utilizados en este estudio no contaban con un control de un segundo análisis, como lo es la técnica de hibridación fluorescente *in situ* (FISH) [19, 85], no fue posible verificar ni realizar un control de las clasificaciones equivocadas 2+. Estudios futuros deberán enfocarse en establecer un *ground truth*¹ [87] y corroborar las principales diferencias en el diagnóstico entre patólogos (2+) y mediante los algoritmos utilizados en HER2 (0/1+), utilizando casos comprobado por diagnóstico.

5.4 CUANTIFICACIÓN AUTOMATIZADA Y VARIABILIDAD EN RECEPTORES HORMONALES

Con respecto a los receptores hormonales, tanto para RE como para RP la metodología desarrollada en la presente tesis, siguió en parte el proceso de evaluación (estimación visual) que realiza el patólogo al comparar la muestra en estudio con la muestra control para asignar un porcentaje de positividad. Esta metodología se basó en la obtención de un coeficiente de positividad que no solo permitió la obtención de porcentajes de positividad, sino también, de ver la

¹ https://en.wikipedia.org/wiki/Ground_truth. Accedido el: 16-11-2016

fluctuación y/o heterogeneidades de las muestras analizadas como las de las muestras control consideradas con 100% de positividad.

Los coeficientes de positividad procedentes de placas que estadísticamente no pertenecieron a la población normal (*Tukey's Honest Significant Difference Test*) fueron revisados posteriormente por patólogos, quienes efectivamente detectaron la existencia de una calidad heterogénea en la marca IHQ en la muestra (especialmente en 3 placas: RE_0001/2 y RP_0006). Dicha heterogeneidad puede ser atribuida a variaciones en la calidad de procesos de tinción, inherentes a factores tales como la calidad de anticuerpos o procedimientos manuales y heterogeneidades del tejido. En este mismo contexto, un estudio similar al nuestro concluyó que el patrón de expresión de RE es mucho más homogéneo que el patrón que expresa RP [83, 84], por lo que se concluye que las muestras procesadas para RP son mucho más heterogéneas [83, 84].

Los resultados obtenidos permiten sugerir tres vías de acción dentro de la adopción de métodos de cuantificación automatizados.

i) El uso de los datos de la distribución normal para definir un coeficiente de positividad para controles cp_{100} y posteriores cálculos de coeficientes de positividad np_{pos} dentro de las placas.

ii) El uso de los datos de la distribución normal para detectar tinciones fuera de los parámetros de normalidad (valores de coeficientes de positividad que no pertenecen a la distribución normal). En la práctica, estas placas procesadas en IHQ se deberían rechazar. Se sugiere revisar el procedimiento de tinción IHQ y el estado de los componentes activos y realizar una segunda tinción.

iii) El permitir variaciones en la densidad de la tinción o marca dentro de un rango razonable y utilizar los valores derivados dentro de estas placas para definir un valor cp_{100} que dé cuenta de las variaciones observadas (se debe definir "rango tolerable o de tolerancia" en el futuro mediante un diseño experimental que considere $N > 10$, probablemente en el rango de $N = 100$ para una posible aplicación clínica).

En RE, los resultados demuestran que existe una alta concordancia tanto inter-observador como al comparar a patólogos versus el algoritmo RG ($K\text{-alpha} > 0.9$,

Capítulo 4, Sección 4.4.3). Aunque estos resultados muestran una tendencia, para ser demostrada en una publicación indexada, se requerirá ampliar el número de placas analizadas a un tamaño muestral cercano a 30 o más.

El hecho que en la medición realizada en forma posterior al proceso de evaluación visual y el análisis con el algoritmo RG en RP se detectaron placas cuyas tinciones mostraron heterogeneidades (RP_0005 y RP_0011), permite apoyar la idea que este algoritmo podría ser utilizado para indicar la existencia de heterogeneidades y la baja precisión de la tinción.

Cabe hacer notar que cuando se trabaja con el promedio de las ROIs obtenidas mediante el algoritmo RG para cada placa virtual versus la evaluación final entregada por los patólogos, el coeficiente de concordancia k-alpha presenta un valor bajo exclusivamente para RP. Este resultado es importante pues permite concluir que los patólogos basaron su evaluación no solo en las ROIs, sino que además consideraron el análisis de toda la placa, o criterios propios de interpretación los cuales no fueron verbalizados.

Nuestro estudio pone en evidencia la variabilidad que existe en la interpretación manual de la IHQ en biopsias de tejido mamario y permite confirmar al mismo tiempo que es posible transformar esta evaluación cualitativa y/o semicuantitativa a una evaluación más cuantitativa y obtener valores más objetivos, utilizando herramientas automatizadas previamente estandarizadas. Un futuro trabajo debiera enfocarse a realizar una clasificación del porcentaje de los núcleos positivos por intensidades de tinción, categorizando una clasificación +1 (débil), +2 (moderado) o +3 (fuerte)¹. Nuestros resultados nos permiten afirmar que esta aproximación es perfectamente realizable mediante procesamiento de imágenes, utilizando el mismo algoritmo creado para este análisis y poniendo algunas condicionantes como tener un conteo de núcleos como mínimo de 300 en un área analizada.

Respecto a la variabilidad puesta en evidencia, así como la heterogeneidad de la tinción IHQ tanto en HER2 y receptores hormonales, existen estudios que muestran

¹ <http://www.cap.org/ShowProperty?nodePath=/UCMCon/Contribution%20Folders/WebContent/pdf/cp-breast-biomarker-template-14.pdf>. Accedido el: 16-11-2016

la existencia de heterogeneidad en la calidad de tinción IHQ debido a que su procesamiento está sujeto a múltiples etapas susceptible de errores en mayor o menor grado que pueden afectar la reproducibilidad y fiabilidad de los resultados, produciendo como consecuencia incertezas o diagnósticos equívocos [47]. Esto pone de manifiesto la necesidad de contar con controles de calidad IHQ externos. Cabe mencionar que en Chile no se cuenta con una institución nacional que promueva los controles de calidad externos, como es el caso de organizaciones internacionales tales como *Nordic immunohistochemical Quality Control (NordiQC)*¹, *United Kingdom National External Quality Assessment Service (UK NEQAS)*² y *Canadian Immunohistochemistry Quality Control (CIQC)*³

Los avances actuales en informática y bioinformática junto con la mejora de la tecnología, sin duda abrirán el camino para una mayor adopción de la PD a mediano y largo plazo, similar a lo acontecido con la radiología digital, lo que permitirá a la comunidad de patólogos en su conjunto aprovechar el verdadero potencial de la MV y PD para el cuidado del paciente. La era digital probablemente redefina la práctica de la patología y el papel del patólogo. Espero que esta tesis proporcione un expediente y entregue un marco referencial para ayudar a patólogos, investigadores y profesionales de laboratorio en la introducción y aplicación de la patología digital en nuestro contexto.

¹ <http://www.nordiqc.org/about.php>. Accedido el 02-02-2017

² <http://www.ukneqas.org.uk/content/PageServer.asp?S=517923677&C=1252&Type=G&ID=62>. Accedido el: 02-02-2017

³ <http://cpqa.ca/>. Accedido el: 02-02-2017

6.RESUMEN & CONCLUSIÓN

El presente trabajo de tesis ha permitido dar inicio a la inserción y uso de las herramientas de PD por parte de un grupo de patólogos chilenos quienes, mediante su participación en este proyecto han colaborado con el trabajo necesario para dar los primeros pasos tendientes a promover el uso de esta nueva tecnología en la práctica clínica y la comunidad médica nacional.

El diseño del proceso de validación y la posterior implementación llevada a cabo en esta tesis permitió definir y obtener los parámetros de configuración requeridos para estandarizar los algoritmos pre-definidos del software, como también proponer nuevas metodologías para la cuantificación de los marcadores tumorales HER2, RE y RP, en una patología específica y de alta prevalencia, como es el cáncer de mama.

La presente tesis permitió además establecer detallados procesos y flujos de trabajo, generado con ello evidencia documentada con el fin de proporcionarles validez y consistencia, de acuerdo a especificaciones y atributos de calidad predeterminados.

La metodología propuesta se ajusta a la realidad local y es realizable utilizando el *software Visiopharm* con sus algoritmos predeterminados o bien creados y/o configurados dentro la misma plataforma de trabajo, como el algoritmo propuesto *RG*. El método desarrollado en la tesis permite su adecuación para llevar a cabo análisis similares dentro la misma plataforma del software de diversos marcadores en otras patologías relevantes.

Respecto al flujo de trabajo para la obtención del *gold standard*, este puede ser utilizado para posteriores trabajos en los cuales se requiera de áreas de análisis ya sea para desarrollar nuevos algoritmos y/o para entrenamiento mediante y a través del cual los especialistas participen de forma remota a través de internet, sin limitantes de tiempo, espacio o lugar determinado.

El uso de herramientas de PD como las placas virtuales, visualizadores *on-line* y acceso remoto o a distancia de bases de datos de imágenes virtuales se muestran

altamente prometedores para facilitar y/o promover proyectos relacionados con el aseguramiento de la calidad de los procesos en la patología, así como estudios de variabilidad, que hasta hace poco se veían limitados en cierta medida debido a temas logísticos de acceso a las placas histopatológicas y/o biopsias.

Los valores de concordancia obtenidos en la cuantificación de los diversos marcadores utilizados en nuestro estudio resultaron equivalentes a la evaluación visual realizada por los patólogos, lo que nos permite concluir que el análisis de imagen permite obtener un diagnóstico equivalente al entregado por el patólogo. Dado que nuestro estudio no contó con análisis mediante técnicas de inmunofluorescencia, serán necesario estudios futuros orientados a medir el rendimiento de los algoritmos bajo esta técnica.

7. BIBLIOGRAFÍA

1. Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. *Annu Rev Pathol.* 2013;8:331-359. doi:10.1146/annurev-pathol-011811-120902.
2. Webster JD, Dunstan RW. Whole-Slide Imaging and Automated Image Analysis: Considerations and Opportunities in the Practice of Pathology. *Vet Pathol.* 2013;51(1):211-223. doi:10.1177/0300985813503570.
3. Pantanowitz L. Digital images and the future of digital pathology. *J Pathol Inform.* 2010;1. doi:10.4103/2153-3539.68332.
4. Brüggmann A, Eld M, Lelkaitis G, et al. Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains. *Breast Cancer Res Treat.* 2012;132(1):41-49. doi:10.1007/s10549-011-1514-2.
5. Słodkowska J, Filas V, Buszkiewicz E, et al. Study on breast carcinoma Her2/neu and hormonal receptors status assessed by automated images analysis systems: ACIS III (Dako) and ScanScope (Aperio). *Folia Histochem Cytobiol.* 2010;48(1):19-25. doi:10.2478/v10042-010-0015-1.
6. Gemen B Van. *Digital Pathology.*; 2008. doi:10.1007/978-3-319-08780-1.
7. Dobson L, Conway C, Hanley A, et al. Image analysis as an adjunct to manual HER-2 immunohistochemical review: A diagnostic tool to standardize interpretation. *Histopathology.* 2010;57(1):27-38. doi:10.1111/j.1365-2559.2010.03577.x.
8. Vainer B. Next generation image analysis for immunohistochemistry quantitation.
9. Arihiro K, Oda M, Ogawa K, et al. Comparison of evaluations of hormone receptors in breast carcinoma by image-analysis using three automated immunohistochemical stainings. *Exp Ther Med.* 2010;1(6):927-932. doi:10.3892/etm.2010.142\etm-01-06-0927 [pii].

10. Kårsnäs A, Larsen R, Dahl A. Learning histopathological patterns. *J Pathol Inform.* 2011;2(2):12. doi:10.4103/2153-3539.92033.
11. Wolff AC, Hammond MEH, Schwartz JN, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med.* 2007;131(1):18-43. doi:10.1043/1543-2165(2007)131[18:ASOCCO]2.0.CO;2.
12. Hammond MEH, Hayes DF, Wolff AC, Mangu PB, Temin S. American society of clinical oncology/college of american pathologists' guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Oncol Pract.* 2010;6(4):195-197. doi:10.1200/JOP.777003.
13. Anatomía patológica. http://www.sld.cu/galerias/pdf/sitios/scap/introduccion_a_la_patologia.pdf. Accessed September 7, 2015.
14. Recomendaciones MDE. ANATOMÍA PATOLÓGICA PARA TUMORES. 2013.
15. Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol.* 2007;25(33):5287-5312. doi:10.1200/JCO.2007.14.2364.
16. Laurinavicius A, Laurinaviciene A, Dasevicius D, et al. Digital image analysis in pathology: Benefits and obligation. *Anal Cell Pathol.* 2012;35(2):75-78. doi:10.3233/ACP-2011-0033.
17. Jaén M, Lejeune M, Pons L, et al. Análisis cuantitativo de técnicas inmunohistoquímicas. Mejora de resultados mediante aplicación de software de análisis de imágenes digitales. 7o Congr Virtual Hispanoam Anatomía Patológica. 2005:1-8. http://www.conganat.org/7congreso/final/vistaImpresion.asp?id_trabajo=285.

18. Palacios J, Andreu X, Calasanz MJ, et al. Recomendación para la determinación de HER2 en cáncer de mama. Consenso nacional de la Sociedad Española de Anatomía Patológica (SEAP) y de la Sociedad Española de Oncología Médica (SEOM). *Rev Española Patol.* 2009;42(1):3-16. doi:10.1016/S1699-8855(09)70147-3.
19. Wolff AC, Hammond MEH, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol.* 2013;31(31):3997-4013. doi:10.1200/JCO.2013.50.9984.
20. Contreras-melendez L, Piottante-becker A, Contreras-seitz M. Variabilidad e la determinación del estado de HER2 por inmunohistoquímica en Chile. 2012;46(1).
21. Nelson D, Ziv a., Bandali KS. Going glass to digital: virtual microscopy as a simulation-based revolution in pathology and laboratory science. *J Clin Pathol.* 2012;65(10):877-881. doi:10.1136/jclinpath-2012-200665.
22. Stathonikos N, Veta M, Huisman A, van Diest PJ. Going fully digital: Perspective of a Dutch academic pathology lab. *J Pathol Inform.* 2013; 4:15. doi:10.4103/2153-3539.114206.
23. Mercan E, Aksoy S, Shapiro LG, Weaver DL, Brunye T, Elmore JG. Localization of Diagnostically Relevant Regions of Interest in Whole Slide Images. 2014 22nd Int Conf Pattern Recognit. 2014:1179-1184. doi:10.1109/ICPR.2014.212.
24. Rojo MG, Punys V, Slodkowska J, Schrader T, Daniel C, Blobel B. Digital pathology in Europe: coordinating patient care and research efforts. *Stud Health Technol Inform.* 2009; 150: 997-1001. doi:10.3233/978-1-60750-044-5-997.
25. Bolton KL, Garcia-Closas M, Pfeiffer RM, et al. Assessment of Automated Image Analysis of Breast Cancer Tissue Microarrays for Epidemiologic

- Studies. *Cancer Epidemiol Biomarkers Prev.* 2010;19(4):992-999. doi: 10.1158/1055-9965.EPI-09-1023.
26. Congress E, Congress I. 11th 5th European Congress on Telepathology and International Congress on Virtual Microscopy. 2012;8(June):6-9.
 27. Tsuchihashi Y. Expanding application of digital pathology in Japan--from education, telepathology to autodiagnosis. *Diagn Pathol.* 2011;6 Suppl 1(Suppl 1):S19. doi:10.1186/1746-1596-6-S1-S19.
 28. OMS Organización Mundial de la Salud. Cáncer de mama en las Américas. 2012;2. doi:10.1017/CBO9781107415324.004.
 29. Ministerio de Salud. Guía Clínica. Cáncer de Mama.; 2015.
 30. Serra C I, Maya G D, Aguayo S C. Cáncer de mama en Chile. Mortalidades comparadas: Niveles nacional, regional y comunal, 1990-2011. Análisis programático. *Rev Chil Cirugía.* 2014;66(2):118-126. doi:10.4067/S0718-40262014000200005.
 31. Prieto M. Epidemiología del cáncer de mama en Chile Epidemiology of breast cancer in Chile. *Rev Med Clin Condes.* 2011;22(4):428-435.
 32. Carruthers C, Carruthers C. *Cancer Research.* Vol 27.; 2000.
 33. Ministerio de Salud. Guía clínica cáncer de mama en personas de 15 años y mas 2005. Minsal. 2005:1-43.
 34. Foundation CR. Su Guía para el Informe de Patología de Cáncer de Seno Su Guía para el Informe de Patología de Cáncer de Seno.
 35. Summary of ASCO / CAP ER and PgR Guideline Recommendations Summary of ASCO / CAP ER and PgR Guideline Recommendations. 2010:2010-2011.
 36. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol.* 2014;5(3):412-424. doi:10.5306/wjco.v5.i3.412.

37. Van Poznak C, Somerfield MR, Bast RC, et al. Use of Biomarkers to Guide Decisions on Systemic Therapy for Women With Metastatic Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol*. 2015;33(24):2695-2704. doi:10.1200/JCO.2015.61.1459.
38. Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. *Mod Pathol*. 2008;21 Suppl 2:S8-S15. doi:10.1038/modpathol.2008.34.
39. Tai W, Mahato R, Cheng K. The role of HER2 in cancer therapy and targeted drug delivery. *J Control Release*. 2010;146(3):264-275. doi:10.1016/j.jconrel.2010.04.009.
40. Cagle PT, Sholl LM, Lindeman NI, et al. Template for reporting results of biomarker testing of specimens from patients with non-small cell carcinoma of the lung. *Arch Pathol Lab Med*. 2014;138(2):171-174. doi:10.5858/arpa.2013-0232-CP.
41. Lange H. Digital Pathology: A Regulatory Overview. *Lab Med*. 2011;42(10):587-591. doi:10.1309/LMA2M9NQQF0ZCVHC.
42. Olympus YK, Long RE, River C, Tillman D. Validation of Digital Pathology In a Healthcare Environment. 2011.
43. Ion ED. Interpretation Manual Breast Cancer.
44. Burstein HJ, Temin S, Anderson H, et al. Adjuvant Endocrine Therapy for Women With Hormone Receptor-Positive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. *J Clin Oncol*. 2014;32(21):2255-2269. doi:10.1200/JCO.2013.54.2258.
45. Carlson RW, Anderson BO, Chopra R, et al. Treatment of breast cancer in countries with limited resources. *Breast J*. 2003;9 Suppl 2:S67-74. doi:10.1046/j.1524-4741.9.s2.6.x.
46. Terrenato I, Arena V, Pizzamiglio S, et al. External Quality Assessment (EQA) program for the preanalytical and analytical immunohistochemical

- determination of HER2 in breast cancer : an experience on a regional scale. 2013:1-7.
47. Vyberg M, Torlakovic E, Seidal T, Risberg B, Nielsen S. Nordic Immunohistochemical Quality Control. :368-371.
 48. Vyberg M. External quality assurance of malignant mesothelioma analyses.
 49. Wasielewski R von, Hasselmann S, Rüschoff J, Fisseler-Eckhoff A, Kreipe H. Proficiency testing of immunohistochemical biomarker assays in breast cancer. *Virchows Arch.* 2008;453(6):537-543. doi:10.1007/s00428-008-0688-4.
 50. Vyberg M, Nielsen S. Proficiency testing in immunohistochemistry---experiences from Nordic Immunohistochemical Quality Control (NordiQC). *Virchows Arch.* 2016;468(1):19-29. doi:10.1007/s00428-015-1829-1.
 51. Miller K. External quality assessment of immunocytochemistry : The experience of the UK National External Quality Assessment Scheme. 1998:360-361.
 52. Congress A, Centre C. 2 nd NordiQC Conference on Applied Immunohistochemistry. 2015:12-13.
 53. Daniel C, Rojo MG, Klossa J, et al. Standardizing the use of whole slide images in digital pathology. *Comput Med Imaging Graph.* 2011;35(7-8):496-505. doi:10.1016/j.compmedimag.2010.12.004.
 54. Ameisen D, Deroulers C, Perrier V, Bouhidel F. Automatic Image Quality Assessment in Digital Pathology : From Idea to Implementation. :148-157.
 55. Deroulers C, Ameisen D, Badoual M, Gerin C, Granier A, Lartaud M. Analyzing huge pathology images with open source software. *Diagn Pathol.* 2013;8:92. doi:10.1186/1746-1596-8-92.
 56. Cooper LAD, Carter AB, Farris AB, et al. Digital Pathology: Data-Intensive Frontier in Medical Imaging: Health-information sharing, specifically of digital pathology, is the subject of this paper which discusses how sharing the rich images in pathology can stretch the capabilities of all otherwi. *Proc IEEE Inst*

- Electr Electron Eng. 2012;100(4):991-1003.
doi:10.1109/JPROC.2011.2182074.
57. NDP . analyze provides a range of software tools for the quantitative analysis of nuclear , cytoplasmic and membrane biomarkers . Nuclei analysis : Ki-67 , ER , PgR , p53 , etc .
 58. Visiopharm. HER2- CONNECT TM.
 59. Jung C, Kim C. Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization. IEEE Trans Biomed Eng. 2010;57(10 PART 2):2600-2604. doi:10.1109/TBME.2010.2060336.
 60. Kourou K, Exarchos TP, Exarchos KP, Karamouzis M V., Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J. 2015;13:8-17. doi:10.1016/j.csbj.2014.11.005.
 61. Moisen GG. Classification and Regression Trees. Encycl Ecol. 2008;(2000):582-588.
doi:http://www.fs.fed.us/rm/pubs_other/rmrs_2008_moisen_g001.pdf.
 62. Morgan J. Classification and Regression Tree Analysis. BuEdu. 2014;(1):16.
<http://www.bu.edu/sph/files/2014/05/MorganCART.pdf>.
 63. Gould J. Classification and Regression Trees (CART) Documentation. 1984.
 64. Inteligencia D De, Informática ETSDI, Nacional U, Educación D. Resumen Este tutorial presenta una introducción al mundo de las máquinas de vectores soporte Support Vector Machine 2 . SVM para clasi cación binaria de ejemplos separables lineal- mente. 2014.
 65. Taeyoonleesascom TYL, Youngmeseociticom YS. MACHINE LEARNING AND BREAST CANCER PREDICTION USING SEER DATA 1973-2012 Machine Learning – Data Mining – Big Data Analytics – Data Scientist Breast Cancer Prediction and Prognosis Machine Learning Methods Comparison of Machine Learning methods Summary and. 2015.

66. Platt JC. Fast training of support vector machines using sequential minimal optimization. *Adv kernel methods*. 1999;376. doi:10.1109/ISKE.2008.4731075.
67. González CA. SVM: Máquinas de Vectores Soporte Contenido.
68. Urmaliya A, Singhai J. Sequential minimal optimization for support vector machine with feature selection in breast cancer diagnosis. 2013 IEEE 2nd Int Conf Image Inf Process IEEE ICIIIP 2013. 2013:481-486. doi:10.1109/ICIIIP.2013.6707638.
69. Müller R, Rogge-Solti A. BPMN for healthcare processes. *CEUR Workshop Proc*. 2011; 705: 65
70. Rojo MG, Rolón E, Calahorra L, et al. Implementation of the Business Process Modelling Notation (BPMN) in the modelling of anatomic pathology processes. *Diagn Pathol*. 2008;3 Suppl 1: S22. doi:10.1186/1746-1596-3-S1-S22.
71. Workshop I, Healthcare U, Technologies S, It ST. Applying Business Process Modeling Notation (BPMN) in Healthcare. 2010.
72. Gwet KL. Agreement coefficients for nominal ratings: A review. *Handb inter-rater Reliab* (4 ed). 2014:27-70.
73. Inter-rater Agreement for Nominal / Categorical Ratings Note that ipsom lorem dummy text generated for this example, so all coding is fictitious. Coding Options: 1 = Positive statement 2 = Negative statement 3 = Neutral statement Total number of all sc. :1-34.
74. L JC, P LVDEL. Evaluación de la concordancia inter-observador en investigación pediátrica: Coeficiente de Kappa. 2008;79(1):54-58.
75. Krippendorff K. Computing Krippendorff ' s Alpha-Reliability. *Dep Pap*. 2011:12. http://repository.upenn.edu/asc_papers.
76. Antoine J-Y, Villaneau J, Lefevre A. Weighted Krippendorff's alpha is a more reliable metrics for multi-coders ordinal annotations: experimental studies on emotion, opinion and coreference annotation. *Proc 14th Conf Eur Chapter*

Assoc Comput Linguist. 2014;(1):550-559.
<http://www.aclweb.org/anthology/E14-1058>.

77. Krippendorff K. Testing the reliability of content analysis data: what is involved and why. *Content Anal Read*. 2007.
78. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: A guide for non-statisticians. *Int J Endocrinol Metab*. 2012;10(2):486-489. doi:10.5812/ijem.3505.
79. Ropella KM. *Introduction to Statistics for Biomedical Engineers*. Vol 2.; 2007. doi:10.2200/S00095ED1V01Y200708BME014.
80. Stevens. Post Hoc Tests in ANOVA. 1999:1-4.
81. Laurinaviciene A, Dasevicius D, Ostapenko V, Jarmalaite S, Lazutka J, Laurinavicius A. Membrane connectivity estimated by digital image analysis of HER2 immunohistochemistry is concordant with visual scoring and fluorescence in situ hybridization results: algorithm evaluation on breast cancer tissue microarrays. *Diagn Pathol*. 2011;6(1):87. doi:10.1186/1746-1596-6-87.
82. Pantanowitz L, Sinard JH, Henricks WH, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med*. 2013;137(12):1710-1722. doi:10.5858/arpa.2013-0093-CP.
83. Rexhepaj E, Brennan DJ, Holloway P, et al. Novel image analysis approach for quantifying expression of nuclear proteins assessed by immunohistochemistry: application to measurement of oestrogen and progesterone receptor levels in breast cancer. *Breast Cancer Res*. 2008;10 (5):R89. doi:10.1186/bcr2187.
84. Nadji M, Gomez-Fernandez C, Ganjei-Azar P, Morales AR. Immunohistochemistry of Estrogen and Progesterone Receptors Reconsidered. *Am J Clin Pathol*. 2004;123:21-27. doi:10.1309/4WV79N2GHJ3X1841.

85. Turashvili G, Leung S, Turbin D, et al. Inter-observer reproducibility of HER2 immunohistochemical assessment and concordance with fluorescent in situ hybridization (FISH): pathologist assessment compared to quantitative image analysis. *BMC Cancer*. 2009;9(1):165. doi:10.1186/1471-2407-9-165.
86. Masmoudi H, Hewitt SM, Petrick N, Myers KJ, Gavrielides MA: Automated quantitative assessment of HER-2/neu immunohistochemical expression in breast cancer. *IEEE Trans Med Imaging* 2009, 28(6):916-925
87. Read J., Pfahringer B., Holmes G. y Frank E., Classifier Chains for Multi-label Classification. *Machine Learning and Knowledge Discovery in Databases* 5782: 254—269, 2009. doi:10.1007/978-3-642-04174-7_17

8.ANEXOS

A.1 Acta de aprobación del Comité de Ética



UNIVERSIDAD DE CHILE - FACULTAD DE MEDICINA
COMITÉ DE ÉTICA DE INVESTIGACIÓN EN SERES



24 ENE. 2017

ACTA DE APROBACIÓN DE PROYECTO

FECHA: 24 de Enero de 2017.

PROYECTO: "DISEÑO DE UN PROCESO DE VALIDACIÓN DE PATOLOGÍA DIGITAL EN CÁNCER DE MAMA".

INVESTIGADOR RESPONSABLE: SRTA. JIMENA LOPEZ

INSTITUCIÓN: PROGRAMA DE ANATOMÍA Y BIOLOGÍA DEL DESARROLLO ICBM, FACULTAD DE MEDICINA, UNIVERSIDAD DE CHILE.

Con fecha 24 de Enero de 2017, el proyecto ha sido analizado a la luz de los postulados de la Declaración de Helsinki, de la Guía Internacional de Ética para la Investigación Biomédica que involucra sujetos humanos CIOMS 1992, y de las Guías de Buena Práctica Clínica de ICH 1996.

Sobre la base de la información proporcionada en el texto del proyecto el Comité de Ética de Investigación en Seres Humanos de la Facultad de Medicina de la Universidad de Chile, estima que el estudio propuesto está bien justificado y que no significa para los sujetos involucrados riesgos físicos, psíquicos o sociales mayores que mínimos.

En virtud de las consideraciones anteriores el Comité otorga la aprobación ética para la realización del estudio propuesto, dentro de las especificaciones del protocolo.

Se extiende este documento por el periodo de un año a contar desde la fecha de aprobación prorrogable según informe de avance y seguimiento bioético.

LUGAR DE REALIZACIÓN DEL ESTUDIO:

- LABORATORIO DE ANÁLISIS DE IMÁGENES CIENTÍFICAS SCIAN – LAB.

Teléfono: 29789536 - Email: comiteceish@med.uchile.cl



UNIVERSIDAD DE CHILE - FACULTAD DE MEDICINA
COMITÉ DE ÉTICA DE INVESTIGACIÓN EN SERES



24 ENE. 2017

INTEGRANTES DEL COMITÉ DE ÉTICA DE INVESTIGACIÓN EN SERES HUMANOS

NOMBRE	CARGO	RELACIÓN CON LA INSTITUCIÓN
Dr. Manuel Oyarzún	Presidente	Sí
Dr. Hugo Amigo	Miembro	Sí
Dra. Lucía Cifuentes	Miembro	Sí
Dra. Grisel Orellana	Miembro	Sí
Sra. Gina Raineri	Miembro	Sí
Dra. María Ángela Delucchi	Miembro	Sí
Dr. Miguel O'Ryan	Miembro	Sí
Sra. Claudia Marshall	Miembro	Sí
Dra. Julieta Gonzalez	Miembro	Sí

Santiago, 24 de Enero de 2017.

Prof. Gina Raineri B.
Secretaria Ejecutiva CEISH

c.c.: - Proyecto Nº 200-2016
- Archivo Acta nº 142

Teléfono: 29789536 - Email: comiteceish@med.uchile.cl

A.2 Listado de casos seleccionados provenientes del HCUCH

Listado HER2					
N°	ID_Institución	ID_CPDAI	Tejido	Tinción	Registro HCUCH?
1		HER2_0001	mama	IHQ	DOS CRUCES 20%
2		HER2_0002	mama	IHQ	UNA CRUZ
3		HER2_0003	mama	IHQ	TRES CRUCES 50%
4		HER2_0004	mama	IHQ	DOS CRUCES 20%
5		HER2_0005	mama	IHQ	DOS CRUCES
6		HER2_0006	mama	IHQ	DOS CRUCES 20%
7		HER2_0007	mama	IHQ	TRES CRUCES 10%
8		HER2_0008	mama	IHQ	DOS CRUCES 90%
9		HER2_0009	mama	IHQ	DOS CRUCES 40%
10		HER2_0010	mama	IHQ	DOS CRUCES 30%
11		HER2_0011	mama	IHQ	DOS CRUCES 30%
12		HER2_0012	mama	IHQ	NEGATIVO
13		HER2_0013	mama	IHQ	NEGATIVO
14		HER2_0014	mama	IHQ	UNA CRUZ 20%
15		HER2_0015	mama	IHQ	TRES CRUCES 100%
16		HER2_0017	mama	IHQ	DOS CRUCES 10%
17		HER2_0018	mama	IHQ	TRES CRUCES 30%
18		HER2_0019	mama	IHQ	TRES CRUCES 40%
19		HER2_0020	mama	IHQ	1+
20		HER2_0021	mama	IHQ	0+
21		HER2_0022	mama	IHQ	1+
22		HER2_0023	mama	IHQ	0+
23		HER2_0024	mama	IHQ	0+
24		HER2_0025	mama	IHQ	1+
25		HER2_0026	mama	IHQ	1+
26		HER2_0027	mama	IHQ	1+

Nota: Selección de casos realizado por patólogo asociado al proyecto FONDEF D11I1096, por razones de confidencialidad, el ID de la institución no es visible.

Listado RE					
N°	ID_Institución	ID_CPDAI	Tejido	Tinción	Registro HCUCH
1		RE_0001	mama	IHQ	90% positivo
2		RE_0002	mama	IHQ	70% positivo
3		RE_0003	mama	IHQ	90% positivo
4		RE_0004	mama	IHQ	80% positivo
5		RE_0005	mama	IHQ	50% positivo
6		RE_0006	mama	IHQ	70% positivo
7		RE_0008	mama	IHQ	90% positivo
8		RE_0009	mama	IHQ	negativo
9		RE_0010	mama	IHQ	90% positivo
10		RE_0011	mama	IHQ	70% positivo

Nota: Selección de casos realizado por patólogo asociado al proyecto FONDEF D1111096, por razones de confidencialidad, el ID de la institución no es visible.

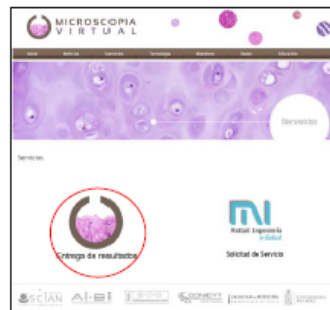
Listado RP					
N°	ID_Institución	ID_CPDAI	Tejido	Tinción	Registro HCUCH ²
1		RP_0001	mama	IHQ-RP	Negativo
2		RP_0002	mama	IHQ	80% positivo
3		RP_0003	mama	IHQ	90% positivo
4		RP_0004	mama	IHQ	60% positivo
5		RP_0005	mama	IHQ	50% positivo
6		RP_0006	mama	IHQ	90% positivo
7		RP_0007	mama	IHQ	Negativo
8		RP_0008	mama	IHQ	70% positivo
9		RP_0009	mama	IHQ	70% positivo
10		RP_0011	mama	IHQ	90% positivo

Nota: Selección de casos realizado por patólogo asociado al proyecto FONDEF D1111096, por razones de confidencialidad, el ID de la institución no es visible.

A.3 Guía rápida de usuarios



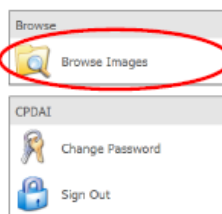
1. Ingrese a www.microscopiavirtual.cl
2. Seleccione ----> Servicios.



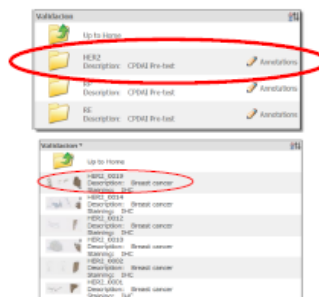
3. Seleccione ----> Entrega de resultados.



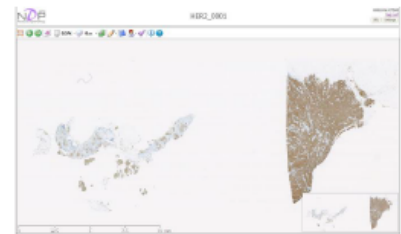
4. Ingrese nombre de usuario y contraseña (Ejemplo: Patologo1, *****).
5. Click ----> Sign in.



6. Ingrese a Browse Images, visualizará una carpeta con las imágenes a analizar.

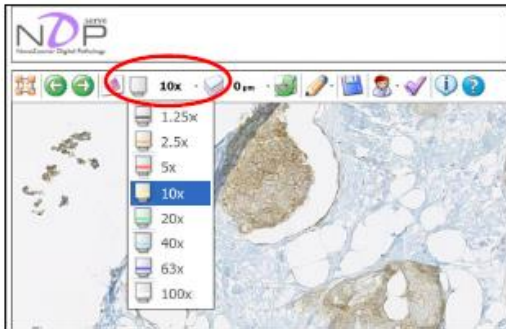


7. Visualizará en pantalla un conjunto de imágenes.
8. Seleccionar la carpeta HER2, seleccione la imagen a analizar



9. Se abrirá la imagen completa, puede navegar en la imagen como el google maps.

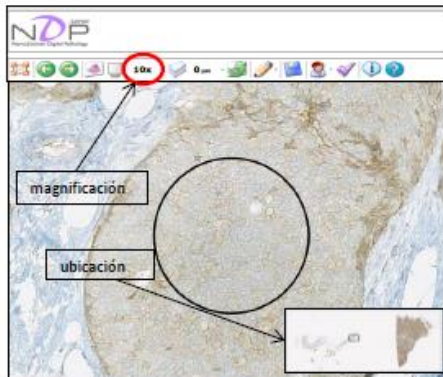
Nota: la imagen de la derecha es control positivo



10. Seleccione el grado de aumento o magnificación de su preferencia.



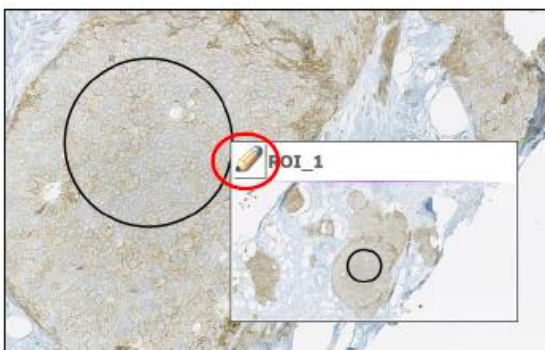
11. Seleccione la opción del círculo para hacer el la region de interés (ROI). Realizar un mínimo de 3 ROI por muestra.



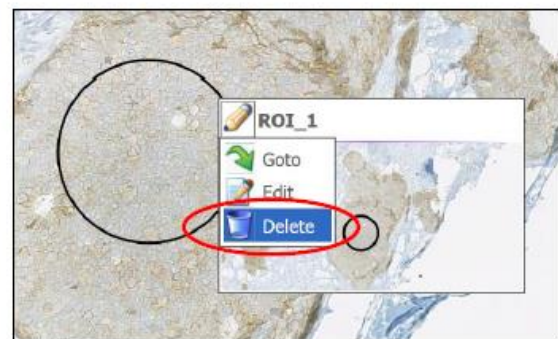
12. Tamaño representativo del ROI a aumento de 20x.



13. Realizado el ROI, enumerarlo y especificar en descripción la clasificación del ROI (0, 1+, 2+, 3+). Realizar esto para cada ROI.



14. En caso que requiera eliminar el ROI, posicione el cursor del mouse en el ROI, se habilitará una opción de edición.



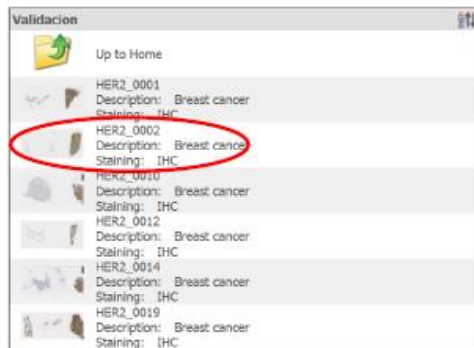
15. Click en el lápiz y seleccione la opción a editar o eliminar.



16. Realizado el ROI vuelve atrás para seleccionar una nueva imagen de la carpeta HER2.



17. Abrir el siguiente link <http://goo.gl/forms/vgCrMo0ahY>
18. Realice la clasificación final para las imágenes de HER2



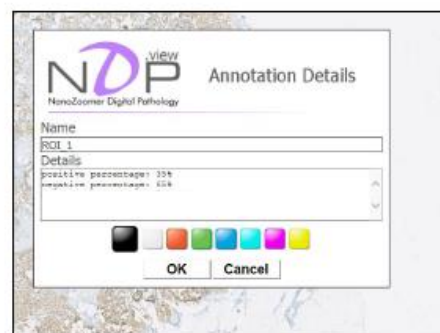
19. Proceder de la misma forma con cada una de las imágenes.
20. Volver a la carpeta principal (Up to Home).



21. Seleccionar la carpeta RP (receptor de progesterona).
22. Proceda de la misma forma que en la carpeta HER2.



23. Realizar la selección de 3 ROIs para cada imagen.



24. Describir la clasificación de análisis como se muestra en la figura. Puede añadir observaciones si es necesario.
25. Proceda de la misma forma con la carpeta RE.



CPDAI - Validación RE&RP

En el siguiente formulario se solicitará realizar la clasificación final del porcentaje de positividad de receptores de hormonas (RE&RP) de cada imagen analizada.

[Visitar](#)

Datos personales

Nombre *

Apellido 1 *

Apellido 2 *

Nombre de la institución de origen *

RP en Cáncer de Mama - Registro de Clasificación

Receptor *

Estrogénico

Progesterona

Observaciones

Historial de respuestas

20%

25. Abrir el siguiente link <http://goo.gl/forms/mvXDNQEO7y>
26. Finalmente realizar la clasificación final para las imágenes de RE (receptor de estrógeno) y RP (receptor de progesterona).



Muchas gracias por su participación!



A.4 Formularios de ingreso de diagnóstico general



MICROSCOPIA V I R T U A L

En el siguiente formulario se solicita realizar la clasificación final (0+,1+,2+,3+), para cada imagen, según código de la imagen.

HER2 en Cancer de Mama - Registro de Clasificación

HER2_0003 *

Ingrese la Clasificación de la imagen

Elige ▼

HER2_0004 *

Ingrese la Clasificación de la imagen

Elige ▼

HER2_0005 *

Ingrese la Clasificación de la imagen

Elige

- 0+ 06 *
Clasificación de la imagen
- 1+
- 2+ 07 *
- 3+ Clasificación de la imagen



MICROSCOPIA V I R T U A L

En el siguiente formulario se solicita realizar la clasificación final del porcentaje de positividad de receptores de hormonas (RE&RP) de cada imagen analizada.

RE en Cancer de mama - Registro de Clasificación

RE_0001

- Positivo
- Negativo

Observaciones

Ingrese porcentaje de positividad

RE_0002

- Positivo
- Negativo

Observaciones

Ingrese porcentaje de positividad

RP en Cancer de Mama - Registro de Clasificación

RP_0001 *

- Positivo
- Negativo

Observaciones

Ingrese porcentaje de positividad

RP_0002 *

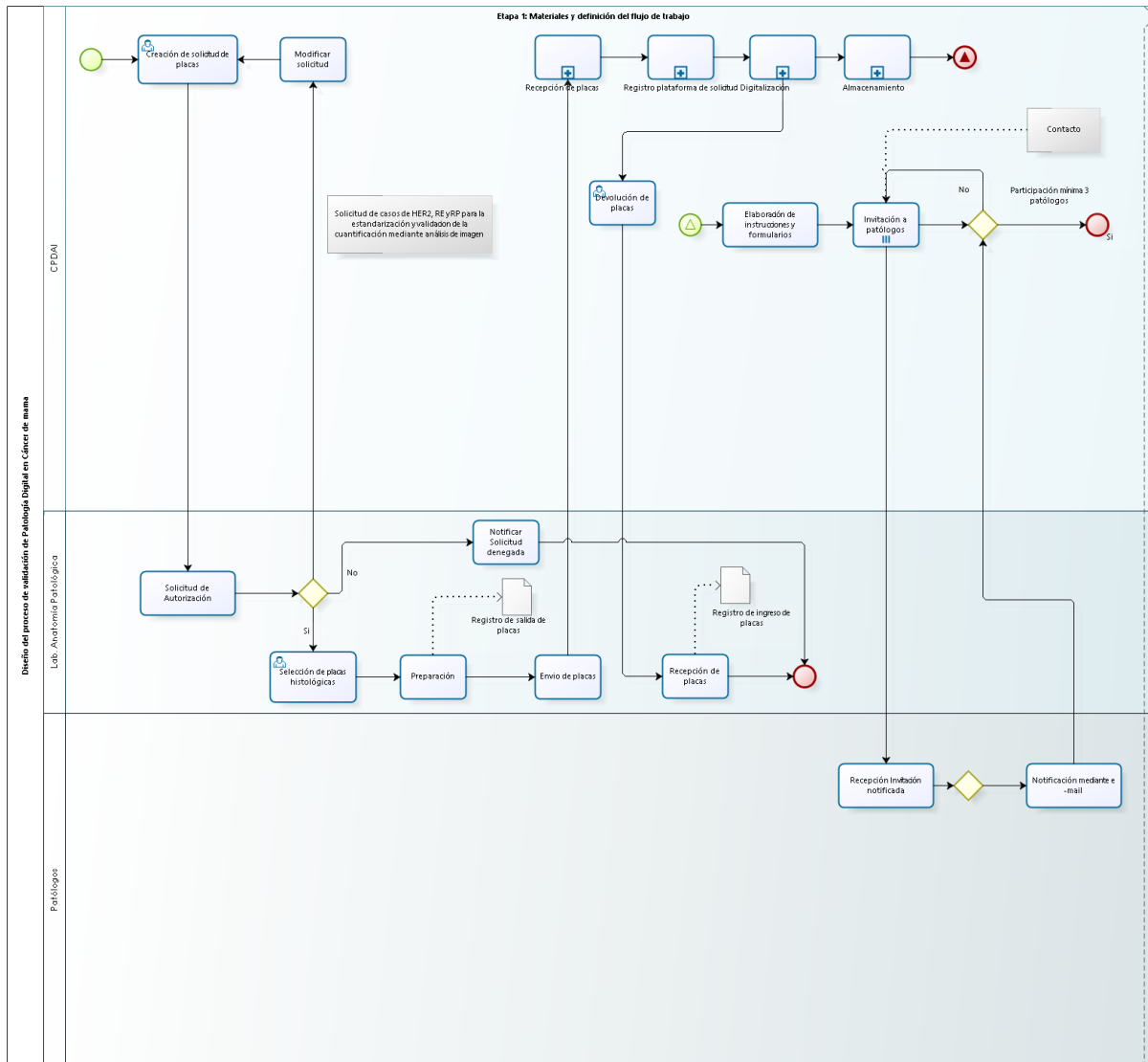
- Positivo
- Negativo

Observaciones

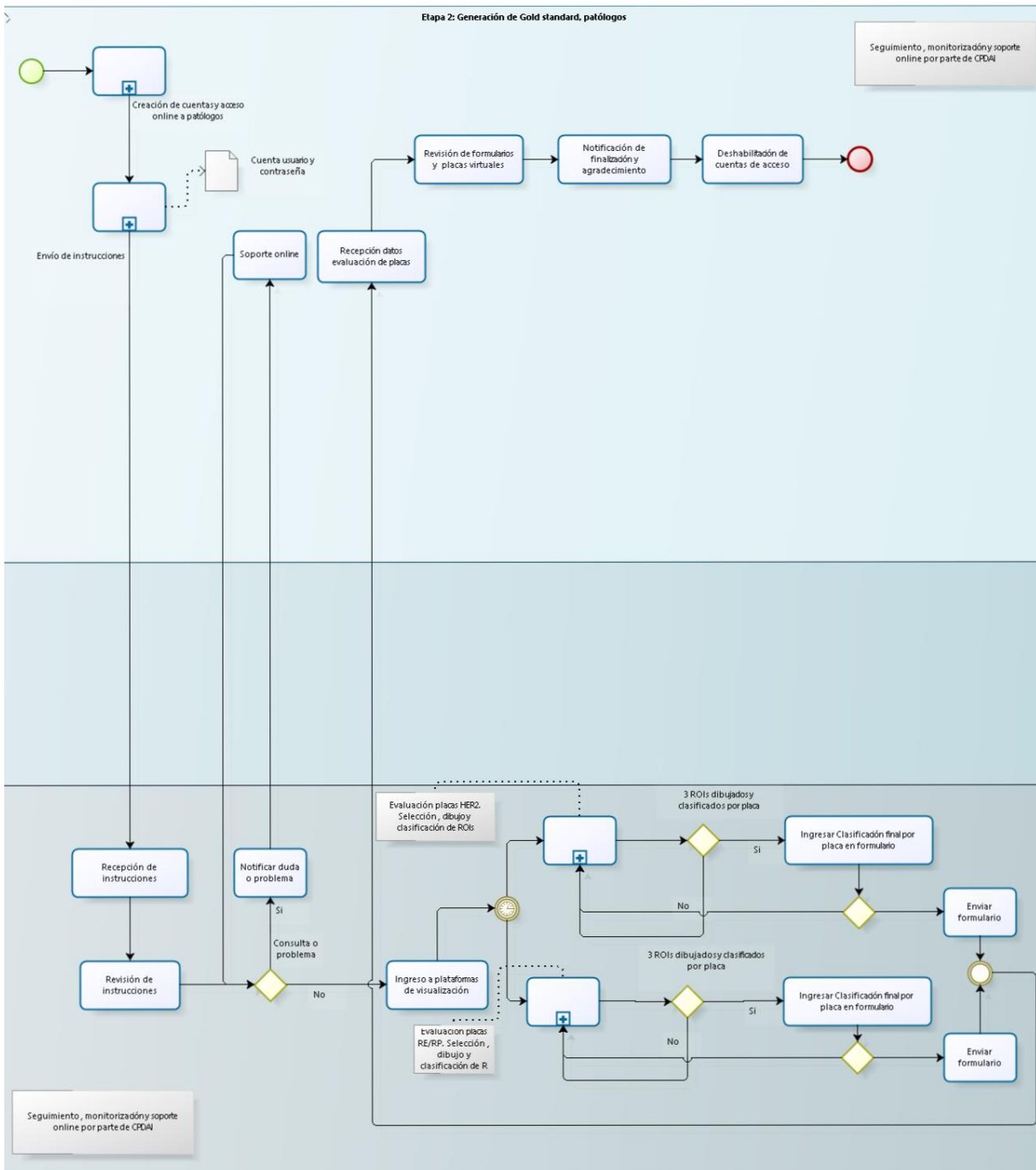
Ingrese porcentaje de positividad

A.5 Diagrama del proceso de validación – Notación BPMN

Etapa 1: Materiales y definición del flujo de trabajo



Etapa 2: Generación del *gold standard* (evaluación visual patólogos)



Etapa 3 y 4: Análisis de imagen (Visiopharm) – Análisis estadístico de concordancia

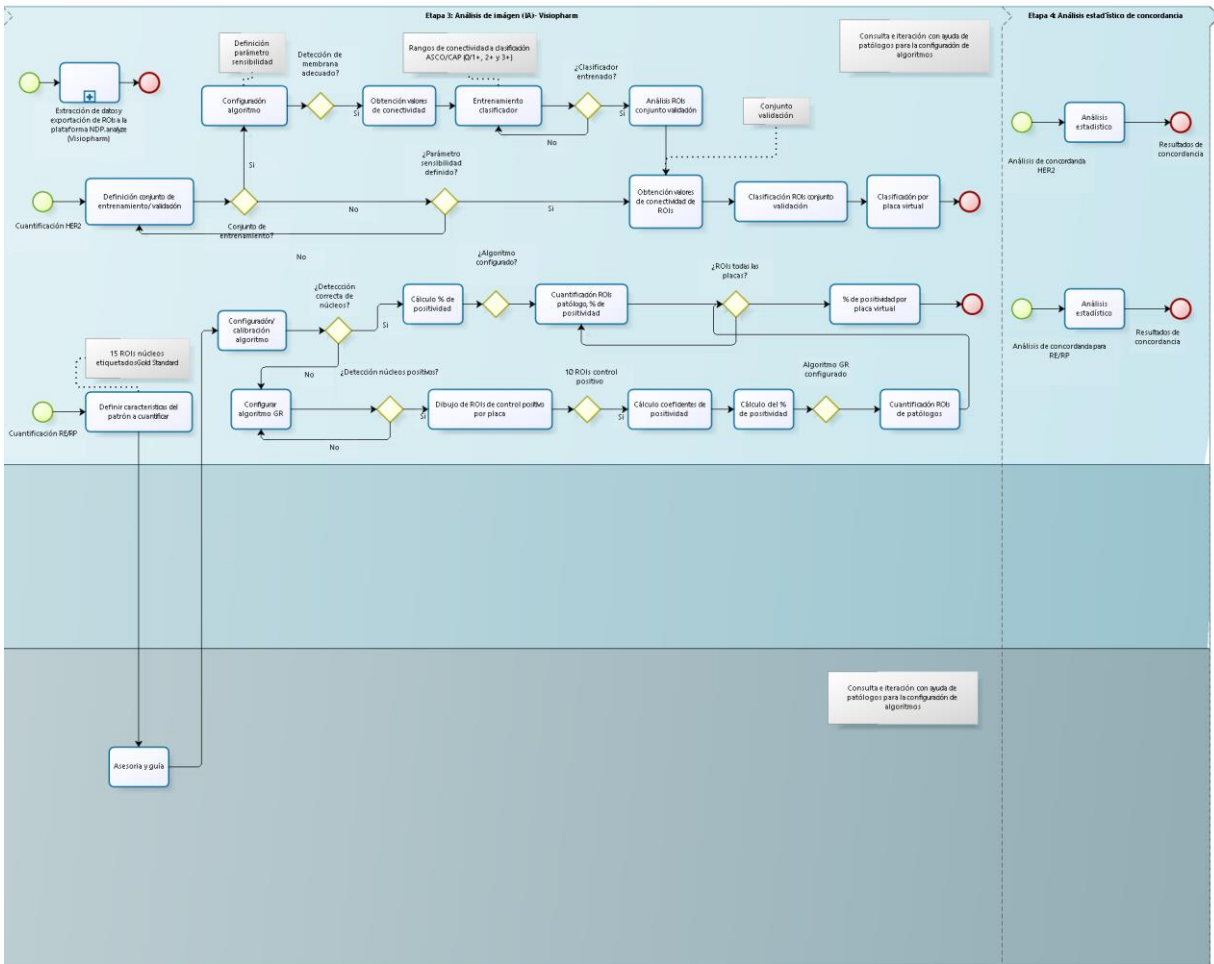
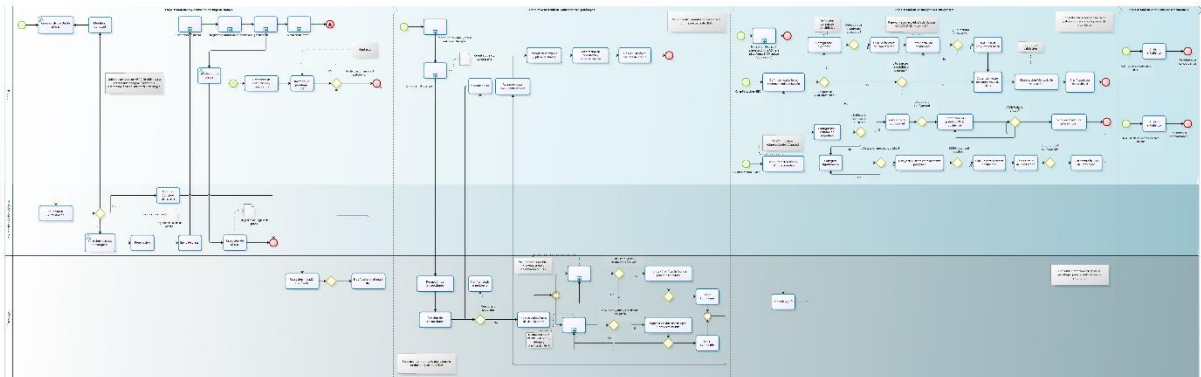


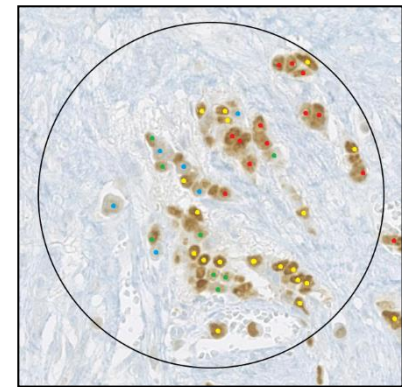
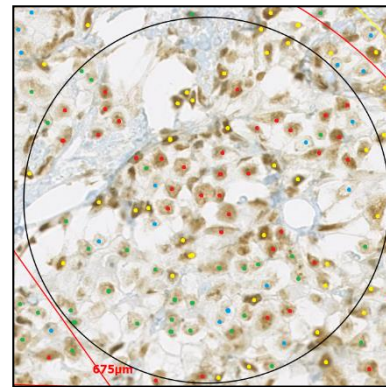
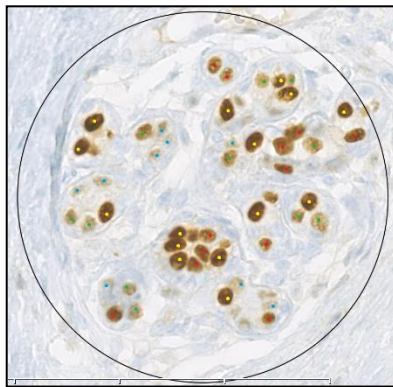
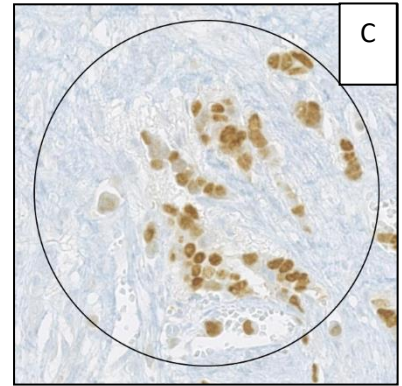
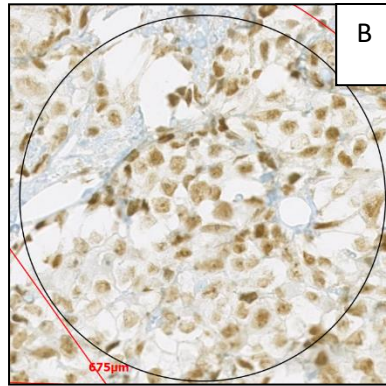
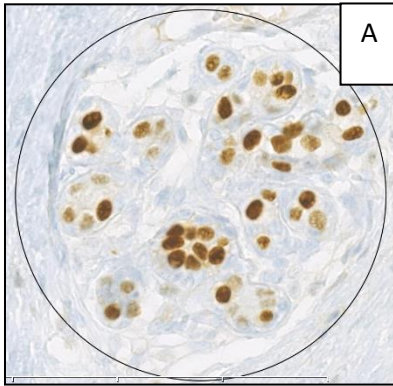
Diagrama del proceso de validación



A.6 Identificación y etiquetado de núcleos en RE

Intensidad de tinción núcleos positivos			Negativo
● Fuerte	● Moderado	● Leve	●

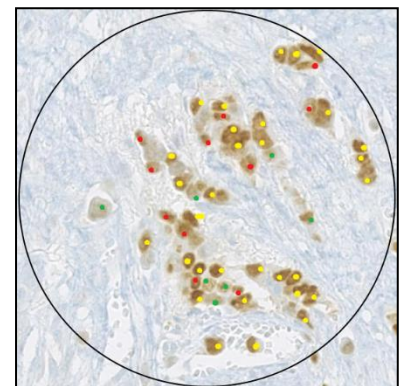
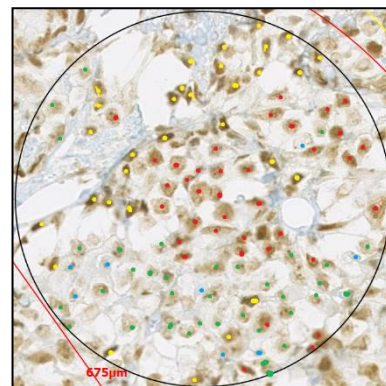
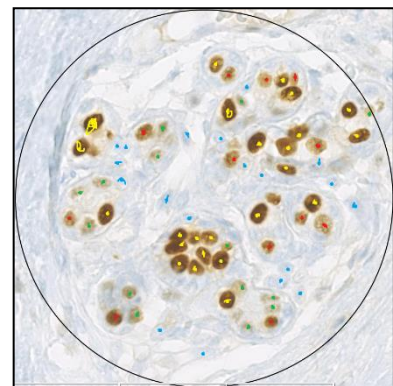
Nota: Para fines de análisis preliminares se considera solo el total de núcleos positivos



	RE GS ROI1		
	Total pos	Negative	Total
Patologo 1	37	9	46

	RE GS ROI2		
	Total pos	Negative	Total
Patologo 1	86	10	96

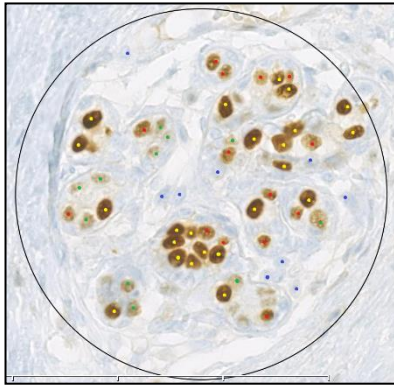
	RE GS ROI3		
	Total pos	Negative	Total
Patologo 1	38	6	44



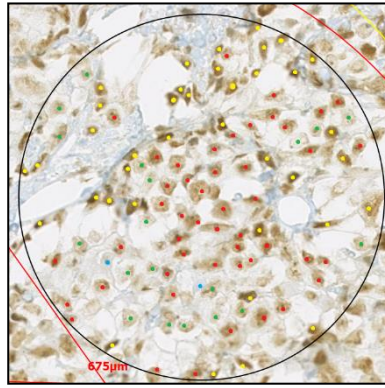
	RE GS ROI1		
	Total pos	Negative	Total
Patologo 2	50	19	69

	RE GS ROI2		
	Total pos	Negative	Total
Patologo 2	104	8	112

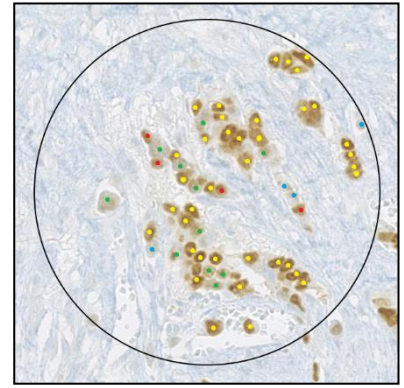
	RE GS ROI3		
	Total pos	Negative	Total
Patologo 2	49	0	49



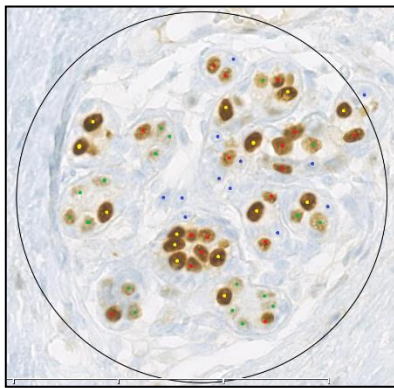
RE GS ROI1			
	Total pos	Negative	Total
Patologo 3	48	9	57



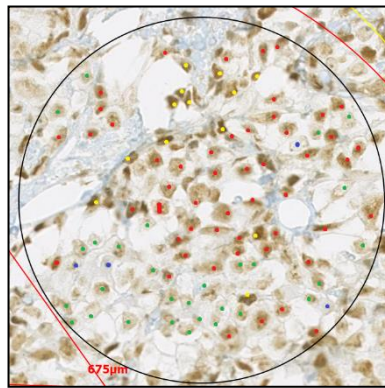
RE GS ROI2			
	Total pos	Negative	Total
Patologo 3	104	2	106



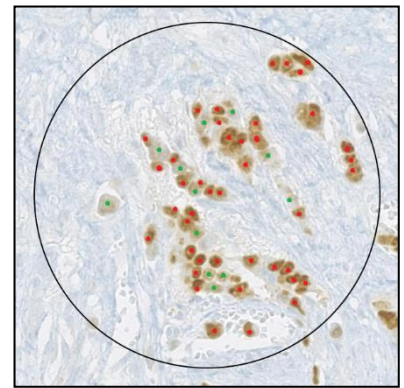
RE GS ROI3			
	Total pos	Negative	Total
Patologo 3	54	4	58



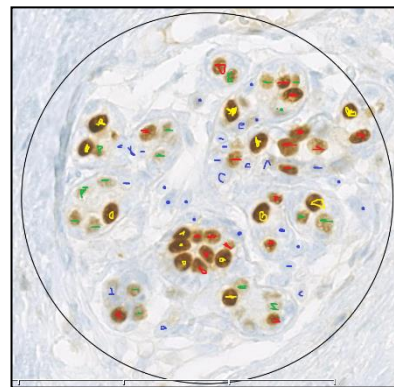
RE GS ROI1			
	Total pos	Negative	Total
Patologo 4	49	11	60



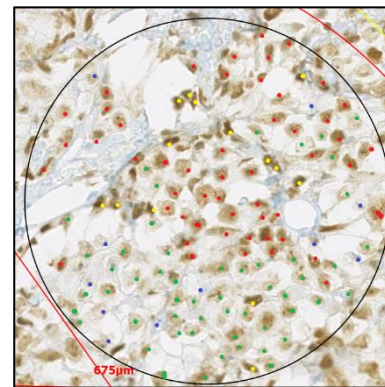
RE GS ROI2			
	Total pos	Negative	Total
Patologo 4	93	4	97



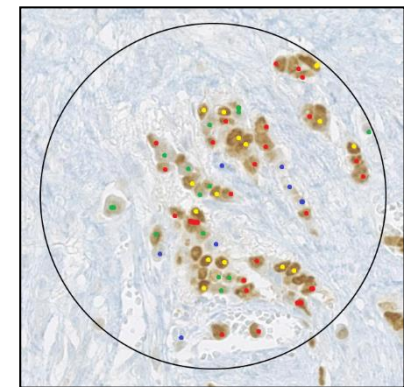
RE GS ROI3			
	Total pos	Negative	Total
Patologo 4	54	0	54



RE GS ROI1			
	Total pos	Negative	Total
Patologo 5	49	28	77



RE GS ROI2			
	Total pos	Negative	Total
Patologo 5	120	11	131



RE GS ROI3			
	Total pos	Negative	Total
Patologo 5	53	7	60

A.7 Resultados de clasificación a 70%, 75% y 80% de sensibilidad – Método CART

Resultados del clasificador entrenado a 70 % de sensibilidad - Método CART

Correctly Classified Instances	90	83.3333 %
Incorrectly Classified Instances	18	16.6667 %
Kappa statistic	0.73	
Mean absolute error	0.1124	
Root mean squared error	0.2774	
Relative absolute error	27.499 %	
Root relative squared error	61.3863 %	
Total Number of Instances	108	

Se observa rendimiento del clasificador de 83.33% correcto. Valor mas bajo comparado a 75% de sensibilidad

Resultados del clasificador entrenado a 75 % de sensibilidad - Método CART

Correctly Classified Instances	99	91.6667 %
Incorrectly Classified Instances	9	8.3333 %
Kappa statistic	0.866	
Mean absolute error	0.0824	
Root mean squared error	0.2187	
Relative absolute error	20.1417 %	
Root relative squared error	48.3981 %	
Total Number of Instances	108	

CART - Training Set				
=== Confusion Matrix ===				
		Image analysis		
		3+	2+	0/1+
Pathologist	3+	36	0	0
	2+	1	15	2
	0/1+	0	6	48

	Correctly Classified
	Incorrectly Classified

Resultados del clasificador entrenado a 80 % de sensibilidad - Método CART

Correctly Classified Instances	95	87.963 %
Incorrectly Classified Instances	13	12.037 %
Kappa statistic	0.8084	
Mean absolute error	0.1151	
Root mean squared error	0.2579	
Relative absolute error	28.1372 %	
Root relative squared error	57.0826 %	
Total Number of Instances	108	

Se observa rendimiento del clasificador de 87.96% correcto. Valor mas bajo comparado a 75% de sensibilidad

Resultados de la clasificación de ROIs del conjunto de validación - Método CART

Correctly Classified Instances 73 61.54 %
 Incorrectly Classified Instances 45 38.46 %
 Kappa statistic 0.3623
 Mean absolute error 0.2758
 Root mean squared error 0.4774
 Relative absolute error 61.29 %
 Root relative squared error 96.40 %
 Total Number of Instances 117

ROIs		ROI_CART			Total
		0/1+	2+	3+	
Evaluación visual (3 Patólogos)	0/1+	39 (33.33%)	12 (10.26%)	0 (0%)	51 (43.59%)
	2+	23 (19.66%)	25 (21.37%)	9 (7.69%)	57 (48.72%)
	3+	0 (0%)	0 (0%)	9 (7.69%)	9 (7.69%)
Total		62 (52.99%)	37 (31.62%)	18 (15.38%)	117 (100%)

A.8 Resultado de traducción de valores de conectividad (75% sensibilidad) mediante arboles de decisión – CART

Traducción de valores de conectividad mediante arboles de decisión - CART									
Nº	Slide ID	Sen 0.75	Pat_ROI	CART_ROI	Final_class_pat			Majority Vote	CART
					P1	P2	P3		
1	HER2_0001P2	0.52888	2+	2+	2+	2+	2+	2+	2+
	HER2_0001P3	0.69127	2+	2+					
	HER2_0001P2	0.65996	2+	2+					
	HER2_0001P3	0.70214	2+	2+					
	HER2_0001P2	0.77354	2+	3+					
	HER2_0001P3	0.80907	2+	3+					
2	HER2_0002P1	0.06555	0/1+	0/1+	0/1+	0/1+	0/1+	0/1+	0/1+
	HER2_0002P2	0.04578	2+	0/1+					
	HER2_0002P3	0.08431	0/1+	0/1+					
	HER2_0002P1	0.03519	0/1+	0/1+					
	HER2_0002P2	0.03838	2+	0/1+					
	HER2_0002P3	0	0/1+	0/1+					
	HER2_0002P1	0.04708	0/1+	0/1+					
	HER2_0002P2	0.30356	2+	2+					
HER2_0002P3	0	0/1+	0/1+						
3	HER2_0006P1	0.82419	2+	3+	2+	2+	0/1+	2+	2+*
	HER2_0006P2	0.7395	2+	3+					
	HER2_0006P3	0.18686	0/1+	2+					
	HER2_0006P1	0.796	2+	3+					
	HER2_0006P2	0.50982	2+	2+					
	HER2_0006P3	0.38749	0/1+	2+					
	HER2_0006P1	0.81908	2+	3+					
	HER2_0006P2	0.4867	2+	2+					
HER2_0006P3	0.12727	0/1+	0/1+						
4	HER2_0008P1	0.77944	2+	3+	0/1+	2+	0/1+	0/1+	2+
	HER2_0008P2	0.43687	0/1+	2+					
	HER2_0008P3	0.16966	0/1+	2+					
	HER2_0008P1	0.05481	0/1+	0/1+					
	HER2_0008P2	0.15137	2+	2+					
	HER2_0008P3	0.25992	0/1+	2+					
	HER2_0008P1	0.07043	0/1+	0/1+					
	HER2_0008P2	0.08739	2+	0/1+					
HER2_0008P3	0.22532	0/1+	2+						

Nº	Slide ID	Sen 0.75	Pat_ROI	CART_ROI	Final_class_pat			Majority Vote	CART
					P1	P2	P3		
5	HER2_0009P1	0.0408	0/1+	0/1+	0/1+	2+	2+	2+	0/1+
	HER2_0009P2	0.2739	2+	2+					
	HER2_0009P3	0.05115	2+	0/1+					
	HER2_0009P1	0.82314	2+	3+					
	HER2_0009P2	0.30598	2+	2+					
	HER2_0009P3	0.05463	2+	0/1+					
	HER2_0009P1	0.10848	0/1+	0/1+					
	HER2_0009P2	0.45175	2+	2+					
	HER2_0009P3	0.10865	2+	0/1+					
6	HER2_0010P1	0	0/1+	0/1+	0/1+	0/1+	0/1+	0/1+	0/1+
	HER2_0010P2	0.0354	0/1+	0/1+					
	HER2_0010P3	0.16155	0/1+	2+					
	HER2_0010P1	0	0/1+	0/1+					
	HER2_0010P2	0.05259	0/1+	0/1+					
	HER2_0010P3	0.11584	0/1+	0/1+					
	HER2_0010P1	0.04579	0/1+	0/1+					
	HER2_0010P2	0.0419	0/1+	0/1+					
	HER2_0010P3	0.11887	0/1+	0/1+					
7	HER2_0011P1	0.05142	0/1+	0/1+	0/1+	2+	2+	2+	0/1+
	HER2_0011P2	0.07968	0/1+	0/1+					
	HER2_0011P3	0.07212	2+	0/1+					
	HER2_0011P1	0.08073	0/1+	0/1+					
	HER2_0011P2	0.08558	0/1+	0/1+					
	HER2_0011P3	0.06445	2+	0/1+					
	HER2_0011P1	0.0521	0/1+	0/1+					
	HER2_0011P2	0.05267	2+	0/1+					
HER2_0011P3	0.08669	2+	0/1+						
8	HER2_0012P1	0.03829	0/1+	0/1+	0/1+	0/1+	0/1+	0/1+	0/1+
	HER2_0012P2	0	0/1+	0/1+					
	HER2_0012P3	0	0/1+	0/1+					
	HER2_0012P1	0.0854	0/1+	0/1+					
	HER2_0012P2	0.02591	0/1+	0/1+					
	HER2_0012P3	0	0/1+	0/1+					
	HER2_0012P2	0.03116	0/1+	0/1+					
	HER2_0012P3	0	0/1+	0/1+					

Nº	Slide ID	Sen 0.75	Pat_ROI	CART_ROI	Final_class_pat			Majority Vote	CART
					P1	P2	P3		
9	HER2_0014P1	0.72905	2+	2+	2+	2+	2+	2+	2+
	HER2_0014P2	0.28989	2+	2+					
	HER2_0014P3	0.64293	2+	2+					
	HER2_0014P1	0.65268	2+	2+					
	HER2_0014P2	0.44043	2+	2+					
	HER2_0014P3	0.72277	2+	2+					
	HER2_0014P1	0.37063	2+	2+,					
	HER2_0014P2	0.48537	2+	2+					
	HER2_0014P3	0.84984	2+	3+					
10	HER2_0017P1	0.3251	0/1+	2+	0/1+	2+	2+	2+	0/1+
	HER2_0017P2	0.10031	2+	0/1+					
	HER2_0017P3	0.07301	2+	0/1+					
	HER2_0017P1	0.34177	0/1+	2+					
	HER2_0017P2	0.30778	2+	2+					
	HER2_0017P3	0.0625	2+	0/1+					
	HER2_0017P1	0.08599	2+	0/1+					
	HER2_0017P2	0.28032	2+	2+					
	HER2_0017P3	0.06173	2+	0/1+					
11	HER2_0019P1	0.73891	3+	3+	3+	3+	3+	3+	3+
	HER2_0019P2	0.88072	3+	3+					
	HER2_0019P3	0.95248	3+	3+					
	HER2_0019P1	0.91435	3+	3+					
	HER2_0019P2	0.93762	3+	3+					
	HER2_0019P3	0.9359	3+	3+					
	HER2_0019P1	0.97451	3+	3+					
	HER2_0019P2	0.96482	3+	3+					
	HER2_0019P3	0.98044	3+	3+					
12	HER2_0020P1	0.06192	0/1+	0/1+	0/1+	2+	2+	2+	0/1+
	HER2_0020P2	0.10454	2+	0/1+					
	HER2_0020P3	0.07642	2+	0/1+					
	HER2_0020P1	0.06265	2+	0/1+					
	HER2_0020P2	0.05469	2+	0/1+					
	HER2_0020P3	0.03636	2+	0/1+					
	HER2_0020P1	0.06094	0/1+	0/1+					
	HER2_0020P2	0.05861	2+	0/1+					
	HER2_0020P3	0.04084	0/1+	0/1+					

Nº	Slide ID	Sen 0.75	Pat_ROI	CART_ROI	Final_class_pat			Majority Vote	CART
					P1	P2	P3		
13	HER2_0025P2	0.12849	2+	0/1+	0/1+	2+	0/1+	0/1+	0/1+
	HER2_0025P3	0.05299	0/1+	0/1+					
	HER2_0025P2	0.45932	2+	2+					
	HER2_0025P3	0.24389	2+	2+					
	HER2_0025P2	0.08052	2+	0/1+					
	HER2_0025P3	0.08185	0/1+	0/1+					
14	HER2_0027P1	0.11641	0/1+	0/1+	0/1+	2+	0/1+	0/1+	2+
	HER2_0027P2	0.15014	2+	2+					
	HER2_0027P3	0.28411	0/1+	2+					
	HER2_0027P2	0.23522	2+	2+					
	HER2_0027P3	0.21553	0/1+	2+					
	HER2_0027P2	0.13814	2+	0/1+					
	HER2_0027P3	0.27549	0/1+	2+					

a) Valor de conectividad obtenidos a 75% de sensibilidad = Sens 0.75

b) La clasificación 0/1+ ambas negativas

c) 9 ROIs fueron excluidas, clasificadas como no evaluables por P1

*Tendiente a una clasificación 3+

Patólogo 1= P1

Patólogo 2 = P2

Patólogo 3 = P3

Voto de mayoría = MV (Majority Vote)

ASCO/CAP	Rango de conectividad
0+	0,00
1+	$0.00 \leq 0.146552$
2+	$0.146552 \leq 0.737601$
3+	$0.737601 \leq 1.00$

A.9 Resultado de traducción de valores de conectividad (60% - 90% sensibilidad) – Método SMO

i) Resultado del entrenamiento del clasificador SMO con datos del conjunto de entrenamiento

SMO - Training Set	
	Results
Correctly Classified Instances (101 ROIs)	93.52%
Incorrectly Classified Instances (7 ROIs)	6.48%
Kappa statistic	0.8947
Mean absolute error	0.2366
Root mean squared error	0.2975
Relative absolute error	57.87%
Root relative squared error	65.83%
Total Number of Instances	108

SMO - Training Set				
=== Confusion Matrix ===				
		Image analysis		
		3+	2+	0/1+
Pathologist	3+	36	0	0
	2+	1	15	2
	0/1+	0	4	50

	Correctly Classified
	Incorrectly Classified

ii) Resultado de clasificación de ROIs del conjunto de validación por el clasificador SMO

	Slide ID	Sen 0.6	Sen 0.65	Sen 0.7	Sen 0.75	Sen 0.80	Sen 0.85	Sen 0.9	Pat_ROI	SMO_ROI	Final_class_pat			Majority Vote	SMO
											P1	P2	P3		
1	HER2_0001IG	0.106169	0.205974	0.452932	0.528879	0.86642	0.949027	0.961575	2+	2+	2+	2+	2+	2+	
	HER2_0001PS	0.104863	0.164358	0.315265	0.691272	0.888683	0.938353	0.945526	2+	2+					
	HER2_0001IG	0.088355	0.142603	0.339503	0.659957	0.894393	0.967751	0.968274	2+	2+					
	HER2_0001PS	0.092507	0.21496	0.452728	0.702141	0.905396	0.956151	0.961472	2+	2+					
	HER2_0001IG	0.136136	0.300586	0.477148	0.77354	0.85825	0.905127	0.937825	2+	2+					
	HER2_0001PS	0.121152	0.203606	0.352582	0.809065	0.924029	0.94957	0.958551	2+	2+					
2	HER2_0002FG	0	0.046292	0.053649	0.065546	0.076834	0.115396	0.143821	1+	1+	0/1+	0/1+	0/1+	0/1+	
	HER2_0002IG	0	0	0.03266	0.045783	0.06354	0.125287	0.343091	2+	1+					
	HER2_0002PS	0	0.05869	0.059062	0.084312	0.129236	0.201231	0.433397	1+	1+					
	HER2_0002FG	0	0	0	0.035189	0.044722	0.061492	0.11207	1+	1+					
	HER2_0002IG	0.025135	0.028922	0.03323	0.038382	0.045209	0.065514	0.095126	2+	1+					
	HER2_0002PS	0	0	0	0	0	0.291864	0.399574	1+	1+					
	HER2_0002FG	0.034436	0.035084	0.039967	0.04708	0.059976	0.073836	0.098321	1+	1+					
	HER2_0002IG	0.070248	0.083297	0.14524	0.303563	0.577089	0.74393	0.846571	2+	2+					
HER2_0002PS	0	0	0	0	0	0	0	1+	1+						
3	HER2_0006FG	0.117049	0.224763	0.498244	0.82419	0.911434	0.943875	0.957643	2+	2+	2+	2+	2+	2+	
	HER2_0006IG	0.124886	0.227419	0.461209	0.739496	0.910655	0.959612	0.976295	2+	2+					
	HER2_0006PS	0.037782	0.043456	0.073948	0.186857	0.644022	0.942342	0.987209	1+	2+					
	HER2_0006FG	0.212297	0.314067	0.641525	0.795997	0.892004	0.955315	0.967015	2+	2+					
	HER2_0006IG	0.073173	0.143204	0.274864	0.509824	0.870055	0.970639	0.983033	2+	2+					
	HER2_0006PS	0.047615	0.071149	0.134416	0.387485	0.765653	0.955141	0.992097	1+	2+					
	HER2_0006FG	0.081699	0.157741	0.27556	0.819082	0.925942	0.950031	0.961212	2+	2+					
	HER2_0006IG	0.142479	0.206731	0.347615	0.486695	0.714543	0.869545	0.940098	2+	2+					
HER2_0006PS	0.030998	0.041425	0.069673	0.127269	0.439729	0.894498	0.989495	1+	2+						

	Slide ID	Sen 0.6	Sen 0.65	Sen 0.7	Sen 0.75	Sen 0.80	Sen 0.85	Sen 0.9	Pat_ROI	SMO_ROI	Final_class_pat			Majority	SMO
											P1	P2	P3	Vote	
11	HER2_0019FG	0.576961	0.653599	0.702015	0.738906	0.762451	0.77005	0.8244	3+	3+	3+	3+	3+	3+	3+
	HER2_0019IG	0.771919	0.811991	0.850581	0.880719	0.901112	0.914061	0.931962	3+	3+					
	HER2_0019PS	0.550316	0.751191	0.88159	0.952481	0.96489	0.970448	0.973328	3+	3+					
	HER2_0019FG	0.767	0.834729	0.879412	0.914355	0.92785	0.944209	0.955812	3+	3+					
	HER2_0019IG	0.760001	0.85779	0.909827	0.937623	0.957413	0.966912	0.978876	3+	3+					
	HER2_0019PS	0.698391	0.833849	0.903895	0.9359	0.948477	0.964641	0.967296	3+	3+					
	HER2_0019FG	0.92982	0.963247	0.968455	0.974515	0.974204	0.976136	0.977087	3+	3+					
HER2_0019IG	0.83544	0.9142	0.944477	0.964824	0.974941	0.984381	0.990904	3+	3+						
HER2_0019PS	0.929638	0.962355	0.974864	0.980442	0.985981	0.98564	0.986464	3+	3+						
12	HER2_0020FG	0.029713	0.03576	0.046555	0.061917	0.10991	0.304667	0.805025	1+	1+	0/1+	2+	2+	2+	0/1+
	HER2_0020IG	0.032335	0.042081	0.063396	0.10454	0.250826	0.610729	0.896389	2+	2+					
	HER2_0020PS	0.033358	0.039329	0.051091	0.076422	0.136712	0.358374	0.763512	2+	1+					
	HER2_0020FG	0.02674	0.032227	0.041188	0.062649	0.104702	0.2731	0.718507	2+	1+					
	HER2_0020IG	0.02539	0.02817	0.040031	0.054691	0.075113	0.185726	0.710486	2+	1+					
	HER2_0020PS	0.0221	0.025432	0.029879	0.036362	0.05944	0.123671	0.541232	2+	1+					
	HER2_0020FG	0.024226	0.03061	0.04135	0.060943	0.097549	0.299285	0.790682	1+	1+					
HER2_0020IG	0.033819	0.039639	0.046616	0.05861	0.106258	0.239447	0.677815	2+	1+						
HER2_0020PS	0.021782	0.02499	0.031724	0.040839	0.065912	0.165062	0.542912	1+	1+						
13	HER2_0025IG	0.047873	0.056344	0.085843	0.128486	0.218833	0.473828	0.802952	2+	1+	0/1+	2+	0/1+	0/1+	0/1+
	HER2_0025PS	0.029804	0.035129	0.043916	0.052989	0.090626	0.256086	0.690738	1+	1+					
	HER2_0025IG	0.090399	0.127422	0.24588	0.459321	0.640482	0.860174	0.951587	2+	2+					
	HER2_0025PS	0.048373	0.066537	0.110012	0.24389	0.472729	0.723469	0.920628	2+	2+					
	HER2_0025IG	0.027663	0.037228	0.049937	0.080517	0.211663	0.416127	0.798784	2+	1+					
HER2_0025PS	0.050953	0.049861	0.054718	0.081845	0.171359	0.359436	0.647198	1+	1+						
14	HER2_0027FG	0.049738	0.068385	0.085164	0.116409	0.173834	0.268329	0.423033	1+	1+	0/1+	2+	0/1+	0/1+	0/1+
	HER2_0027IG	0.067203	0.082377	0.108637	0.150141	0.209962	0.365385	0.60647	2+	1+					
	HER2_0027PS	0.078425	0.125017	0.179354	0.284115	0.414822	0.638098	0.813422	1+	2+					
	HER2_0027IG	0.078759	0.100962	0.151899	0.235222	0.384831	0.583006	0.765515	2+	2+					
	HER2_0027PS	0.071369	0.096052	0.144872	0.215528	0.295326	0.458931	0.649162	1+	1+					
	HER2_0027IG	0.057441	0.075761	0.093303	0.138138	0.224178	0.374161	0.626777	2+	1+					
HER2_0027PS	0.069443	0.1058	0.162194	0.275488	0.411955	0.641192	0.80742	1+	2+						

a) Valor de conectividad obtenidos en el rango de sensibilidad de 60% -90%

b) La clasificación 0/1+ ambas negativas

c) 9 ROIs fueron excluidas, clasificadas como no evaluables por P1

Patólogo 1 = P1

Patólogo 2 = P2

Patólogo 3 = P3

Voto de mayoría = MV (Majority Vote)

A.10 Resultados de cuantificación de receptores de estrógeno

Slide ID	Control	Pat	Pat_pos_ %_ROI	Pat_pos_ _avg_%	RG_ROI	RG_% norm	RG_roun ded_%	RG_avg_r oi_%	RG_rpos_ avg_%	Final_Class_Pat			Pat_avg	RG_avg	RG_final discrete
	Coeficien ↑				Coeficient					P1	P2	P3			
RE_0001	0.63402	P1_1	100		0.509	80.28	80%	75.13	75	90%	100%	100%	97%	83%	80%
		P1_2	70	87	0.441	69.56	70%								
		P1_3	90		0.479	75.55	80%								
		P2_1	100		0.58	91.48	90%	84.22	84						
		P2_2	100	100	0.57	89.90	90%								
		P2_3	100		0.452	71.29	70%	91.69	92						
		P3_1	100		0.541	85.33	90%								
		P3_2	100	100	0.526	82.96	80%								
		P3_3	100		0.677	106.78	100%								
RE_0002	0.55105	P1_1	100		0.68	123.40	100%	120.80	100	100%	100%	80%	93%	100%	100%
		P1_2	100	100	0.677	122.86	100%								
		P1_3	100		0.64	116.14	100%								
		P2_1	100		0.688	124.85	100%	116.26	100						
		P2_2	100	98	0.65	117.96	100%								
		P2_3	95		0.584	105.98	100%	122.01	100						
		P3_1	90		0.72	130.66	100%								
		P3_2	70	77	0.72	130.66	100%								
		P3_3	70		0.577	104.71	100%								
RE_0003	0.37808	P1_1	100		0.614	162.40	100%	167.51	100	100%	100%	100%	100%	100%	100%
		P1_2	100	100	0.663	175.36	100%								
		P1_3	100		0.623	164.78	100%								
		P2_1	100		0.552	146.00	100%	166.54	100						
		P2_2	100	100	0.662	175.10	100%								
		P2_3	100		0.675	178.53	100%	136.74	100						
		P3_1	100		0.485	128.28	100%								
		P3_2	100	100	0.505	133.57	100%								
		P3_3	100		0.561	148.38	100%								
RE_0004	0.38579	P1_1	100		0.514	133.23	100%	140.75	100	100%	90%	100%	97%	100%	100%
		P1_2	100	100	0.559	144.90	100%								
		P1_3	100		0.556	144.12	100%								
		P2_1	100		0.573	148.53	100%	144.73	100						
		P2_2	90	90	0.568	147.23	100%								
		P2_3	80		0.534	138.42	100%	156.30	100						
		P3_1	100		0.667	172.89	100%								
		P3_2	100	100	0.56	145.16	100%								
		P3_3	100		0.582	150.86	100%								
RE_0005	0.33656	P1_1	100		0.671	199.37	100%	199.37	100	100%	100%	100%	100%	100%	100%
		P1_2	Ino evalua	100	Sin info	-	-								
		P1_3	Ino evaluable		Sin info	-	-	-	-						
		P2_1	100		0.718	213.33	100%	207.39	100						
		P2_2	100	100	0.675	200.56	100%								
		P2_3	100		0.701	208.28	100%								
		P3_1	100		0.695	206.50	100%	207.79	100						
		P3_2	100	100	0.74	219.87	100%								
		P3_3	100		0.663	196.99	100%								
RE_0006	0.44929	P1_1	100		0.463	103.05	100%	104.17	100	100%	100%	70%	90%	99%	100%
		P1_2	100	97	0.443	98.60	100%								
		P1_3	90		0.498	110.84	100%								
		P2_1	100		0.505	112.40	100%	107.87	100						
		P2_2	100	100	0.463	103.05	100%								
		P2_3	100		0.486	108.17	100%	96.15	96						
		P3_1	70		0.418	93.04	90%								
		P3_2	70	70	0.447	99.49	100%								
		P3_3	70		0.431	95.93	100%								

Slide ID	Control	Pat	Pat_pos_ %_ROI	Pat_pos_ avg_%	RG_ROI	RG_%_ norm	RG_roun ded_%	RG_avg_r oi_%	RG_rpos_ avg_%	Final_Class_Pat			Pat_avg	RG_avg	RG_final_ discrete
	Coeficien				Coeficient					P1	P2	P3			
RE_0008	0.51492	P1_1	100		0.498	96.71	100%	96.13	96	95%	93%	90%	93%	93%	90%
		P1_2	100	100	0.506	98.27	100%								
		P1_3	100		0.481	93.41	90%								
		P2_1	100		0.432	83.90	80%	92.77	93	95%	93%	90%	93%	93%	90%
		P2_2	100	97	0.521	101.18	100%								
		P2_3	90		0.48	93.22	90%								
		P3_1	100		0.488	94.77	90%	96.84	97	95%	93%	90%	93%	93%	90%
		P3_2	100	100	0.524	101.76	100%								
		P3_3	100		0.484	94.00	90%								
RE_0009	0.41970	P1_1	0		0(n)	0.00	0%	0.00	0	0%	0%	0%	0%	0%	0%
		P1_2	0	0	0(n)	0.00	0%								
		P1_3	0		0(n)	0.00	0%								
		P2_1	0		0(n)	0.00	0%	0.00	0	0%	0%	0%	0%	0%	0%
		P2_2	0	0	0(n)	0.00	0%								
		P2_3	0		0(n)	0.00	0%								
		P3_1	0		0(n)	0.00	0%	0.00	0	0%	0%	0%	0%	0%	0%
		P3_2	0	0	0(n)	0.00	0%								
		P3_3	0		0(n)	0.00	0%								
RE_0010	0.42940	P1_1	100		0.43	100.14	100%	89.82	90	100%	93%	100%	98%	94%	90%
		P1_2	100	100	0.399	92.92	90%								
		P1_3	100		0.328	76.39	80%								
		P2_1	90		0.416	96.88	100%	93.23	93	100%	93%	100%	98%	94%	90%
		P2_2	90	90	0.365	85.00	90%								
		P2_3	90		0.42	97.81	100%								
		P3_1	100		0.559	130.18	100%	105.19	100	100%	93%	100%	98%	94%	90%
		P3_2	90	93	0.376	87.56	90%								
		P3_3	90		0.42	97.81	100%								
RE_0011	0.43092	P1_1	100		0.302	70.08	70%	74.96	75	100%	90%	90%	93%	74%	70%
		P1_2	100	98	0.362	84.01	80%								
		P1_3	95		0.305	70.78	70%								
		P2_1	100		0.363	84.24	90%	72.40	72	100%	90%	90%	93%	74%	70%
		P2_2	90	93	0.299	69.39	70%								
		P2_3	90		0.274	63.59	60%								
		P3_1	90		0.29	67.30	70%	76.19	76	100%	90%	90%	93%	74%	70%
		P3_2	90	90	0.351	81.45	80%								
		P3_3	90		0.344	79.83	80%								

A.11 Resultados de cuantificación de receptores de progesterona

Slide ID	Control	Pat	Pat_pos %_ROI	Pat_pos _avg_%	RG_ROI	RG_%_n	RG_rou	RG_avg_	RG_rpo	Final_Class_Pat			Pat_avg	RG_avg	RG_final _discrete
	Coeficient				Coeficien t	orm	nded_%	roi_%	s_avg_ %	P1	P2	P3			
RP_0001	0.442871	P1_1	0		0.412	92.95	90%			0%	0%	0%	0%	46%	50%
		P1_2	ROI no evaluable		sin info	-	-	92.95	93						
		P1_3	ROI no evaluable		sin info	-	-								
		P2_1	0		0.473	106.73	100%								
		P2_2	0	0	0.000	0.00	0%	58.45	58						
		P2_3	0		0.304	68.63	70%								
		P3_1	0		0.250	56.40	60%								
		P3_2	0	0	0.000	0.00	0%	18.80	19						
		P3_3	0		0.000	0.00	0%								
RP_0002	0.476483	P1_1	90		0.701	147.13	100%			100%	100%	90%	97%	100%	100%
		P1_2	100	97	0.674	141.44	100%	146.10	100						
		P1_3	100		0.713	149.71	100%								
		P2_1	100		0.696	146.10	100%								
		P2_2	100	100	0.658	138.06	100%	130.88	100						
		P2_3	100		0.517	108.49	100%								
		P3_1	90		0.514	107.80	100%								
		P3_2	90	90	0.530	111.27	100%	112.73	100						
		P3_3	90		0.568	119.11	100%								
RP_0003	0.433832	P1_1	40		0.424	97.74	100%			95%	90%	95%	93%	100%	100%
		P1_2	100	80	0.650	149.75	100%	132.34	100						
		P1_3	100		0.649	149.52	100%								
		P2_1	100		0.654	150.76	100%								
		P2_2	80	93	0.641	147.67	100%	151.95	100						
		P2_3	100		0.683	157.40	100%								
		P3_1	95		0.611	140.77	100%								
		P3_2	100	97	0.605	139.48	100%	140.78	100						
		P3_3	95		0.616	142.08	100%								
RP_0004	0.464733	P1_1	100		0.606	130.36	100%			100%	90%	90%	93%	100%	100%
		P1_2	100	100	0.583	125.43	100%	124.56	100						
		P1_3	100		0.548	117.89	100%								
		P2_1	90		0.467	100.48	100%								
		P2_2	90	90	0.506	108.87	100%	104.98	100						
		P2_3	90		0.491	105.60	100%								
		P3_1	90		0.488	104.97	100%								
		P3_2	90	90	0.554	119.26	100%	108.70	100						
		P3_3	90		0.473	101.86	100%								
RP_0005	0.455744	P1_1	90		0.597	130.92	100%			-	95%	0%	48%	83%	80%
		P1_2	ROI no e	90	sin info	-	-	130.92	100						
		P1_3	ROI no evaluable		sin info	-	-								
		P2_1	95		0.665	146.02	100%								
		P2_2	90	95	0.651	142.88	100%	134.02	100						
		P2_3	100		0.516	113.15	100%								
		P3_1	0		0.369	81.05	80%								
		P3_2	0	0	0.000	0.00	0%	65.32	65						
		P3_3	0		0.524	114.90	100%								
RP_0006	0.526008	P1_1	100		0.668	127.01	100%			90%	100%	100%	97%	100%	100%
		P1_2	100	100	0.675	128.24	100%	126.18	100						
		P1_3	100		0.649	123.30	100%								
		P2_1	100		0.556	105.63	100%								
		P2_2	100	100	0.582	110.58	100%	112.24	100						
		P2_3	100		0.634	120.51	100%								
		P3_1	100		0.597	113.44	100%								
		P3_2	100	100	0.606	115.12	100%	114.18	100						
		P3_3	100		0.599	113.97	100%								

Slide ID	Control	Pat	Pat_pos %_ROI	Pat_pos _avg_%	RG_ROI	RG_%_n orm	RG_rou nded_%	RG_avg_ roi_%	RG_rpo s_avg_ %	Final_Class_Pat			Pat_avg	RG_avg	RG_final _discrete
	Coefficient				Coefficien †					P1	P2	P3			
RP_0007	0.447435	P1_1	0		0.000	0.00	0%	-	-	0%	0%	0%	0%	6%	10%
		P1_2	ROI no evaluable		sin info	-	-	-	-						
		P1_3	ROI no evaluable		sin info	-	-	-	-						
		P2_1	0		0.000	0.00	0%	-	-						
		P2_2	0	0	0.000	0.00	0%	0.00	0						
		P2_3	0		0.000	0.00	0%	-	-						
		P3_1	0		0.000	0.00	0%	-	-						
		P3_2	0	0	0.000	0.00	0%	13.36	13						
RP_0008	0.508374	P1_1	0		0.000	0.00	0%	-	-	-	0%	0%	0%	10%	10%
		P1_2	ROI no evaluable		sin info	-	-	0.00	0						
		P1_3	ROI no evaluable		sin info	-	-	-	-						
		P2_1	10		0.378	74.45	70%	-	-						
		P2_2	0	3	0.000	0.00	0%	24.82	25						
		P2_3	0		0.000	0.00	0%	-	-						
		P3_1	0		0.000	0.00	0%	-	-						
		P3_2	0	0	0.000	0.00	0%	0.00	0						
RP_0009	0.465748	P1_1	40		0.426	91.46	90%	-	-	40%	55%	60%	52%	90%	90%
		P1_2	50	40	0.454	97.58	100%	90.11	90						
		P1_3	30		0.379	81.29	80%	-	-						
		P2_1	40		0.312	66.89	70%	-	-						
		P2_2	50	53	0.378	81.23	80%	80.91	81						
		P2_3	70		0.441	94.62	90%	-	-						
		P3_1	60		0.542	116.27	100%	-	-						
		P3_2	60	60	0.548	117.73	100%	116.26	100						
RP_0011	0.485469	P1_1	100		0.623	128.41	100%	-	-	100%	85%	30%	72%	91%	90%
		P1_2	100	100	0.503	103.62	100%	113.65	100						
		P1_3	100		0.529	108.92	100%	-	-						
		P2_1	100		0.514	105.96	100%	-	-						
		P2_2	80	93	0.359	73.93	70%	96.33	96						
		P2_3	100		0.530	109.11	100%	-	-						
		P3_1	60		0.368	75.81	80%	-	-						
		P3_2	30	40	0.487	100.36	100%	83.34	83						
P3_3	30		0.358	73.85	70%	-	-								