

DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

DESEMPEÑO DIAGNÓSTICO Y REPRODUCIBILIDAD INTEROBSERVADOR DEL  
GRADO DE GLEASON EN LA INTERPRETACIÓN HISTOPATOLÓGICA  
CONVENCIONAL Y MEDIANTE EL ANÁLISIS DE IMÁGENES POR SOFTWARE

CARLOS ANDRÉS SÁNCHEZ RUEDA

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE SALUD – ESCUELA DE MEDICINA  
DEPARTAMENTO DE PATOLOGÍA  
ESPECIALIZACION EN PATOLOGÍA  
BUCARAMANGA

2023

DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

DESEMPEÑO DIAGNÓSTICO Y REPRODUCIBILIDAD INTEROBSERVADOR DEL  
GRADO DE GLEASON EN LA INTERPRETACIÓN HISTOPATOLÓGICA  
CONVENCIONAL Y MEDIANTE EL ANÁLISIS DE IMÁGENES POR SOFTWARE

Autor

CARLOS ANDRÉS SÁNCHEZ RUEDA

Trabajo de grado para optar por el título de Especialista en Patología

Director

OLGA MERCEDES ÁLVAREZ OJEDA

MD Patóloga

Codirector

TANIA MENDOZA HERRERA

MD Internista – Epidemióloga

UNIVERSIDAD INDUSTRIAL DE SANTANDER

FACULTAD DE SALUD – ESCUELA DE MEDICINA

DEPARTAMENTO DE PATOLOGÍA

ESPECIALIZACIÓN EN PATOLOGÍA

BUCARAMANGA

2023

**Tabla de Contenido**

Introducción .....	11
1. Justificación .....	12
2. Objetivos .....	14
2.1. Objetivo General .....	15
2.2. Objetivos Específicos.....	15
3. Marco Teórico.....	16
3.1. La Próstata .....	16
3.2. El Adenocarcinoma de Próstata .....	16
3.3. Escala de Gleason .....	17
3.3.1 Gleason Patrones 1 y 2.....	18
3.3.2 Gleason Patrón 3 .....	18
3.3.3 Gleason Patrón 4 .....	19
3.3.4 Gleason Patrón 5 .....	19
3.4. Factores de Riesgo .....	20
3.5. Patología Digital .....	21
3.6. Qupath y Demarcación de Imágenes de Patología Digital .....	23
3.7. Errores y Sesgos de la Evaluación Histopatológica Tradicional .....	24
3.7.1 Sesgos Cognitivos.....	24

## DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

3.7.2	Sesgos Visuales.....	25
3.8.	Análisis Digital de Imágenes .....	26
4.	Diseño Metodológico.....	28
4.1	Diseño del Estudio .....	28
4.2.	Población.....	31
4.2.1	Criterios de Inclusión.....	31
4.2.2	Criterios de Exclusión.....	32
4.3.	Instrumentos del Estudio.....	32
4.4.	Procesamiento de Datos .....	33
5.	Evaluación Crítica del Protocolo .....	33
5.1.	Ventajas del Diseño .....	33
5.2.	Desventajas del Diseño .....	34
6.	Análisis Estadístico de las Variables .....	34
7.	Definición de las Variables.....	35
8.	Cronograma.....	36
9.	Presupuesto .....	37
10.	Consideraciones Éticas .....	37
10.1.	Principios Éticos .....	38
10.1.1	Principio de Beneficencia .....	38

## DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

10.1.2	Principio de Respeto a las Personas .....	39
10.1.3	Principio de Justicia .....	39
10.2.	Tratamiento de los Datos Personales .....	40
10.3.	Clasificación del Riesgo de la Investigación .....	41
10.4.	Procedimientos para Abordar a los Participantes .....	41
10.5.	Potenciales Beneficiarios .....	42
10.6.	Riesgos para los Participantes e Investigadores.....	43
10.7.	Consentimiento Informado .....	43
11.	Resultados .....	43
12.	Discusión y Conclusión .....	51
13.	Conflicto de Intereses .....	53
	Referencias Bibliográficas .....	54
	Apéndices.....	62

**Lista de Figuras**

**Figura 1.** Captura de la ventana de QuPath. .... 23

**Figura 2.** Gráfica de barras que compara el coeficiente Kappa sin ayuda de la predicción y el coeficiente Kappa con ayuda de la predicción generada mediante inteligencia artificial. .... 45

**Figura 3.** Gráfica de barras que compara la sensibilidad sin ayuda de la predicción y la sensibilidad con ayuda de la predicción generada mediante inteligencia artificial. .... 49

**Figura 4.** Gráfica de barras que compara el VPP sin ayuda de la predicción y el VPP con ayuda de la predicción generada mediante inteligencia artificial..... 50

**Lista de Tablas**

<b>Tabla 1.</b> Resumen de los criterios usados para la clasificación mediante el score de Gleason para grados 3, 4 y 5, y ejemplos de imágenes histopatológicas extraídas del Dataset. ....	20
<b>Tabla 2.</b> Características de los sets de imágenes histopatológicas que se usarán en la primera fase del estudio y el tiempo que tardará su demarcación. ....	30
<b>Tabla 3.</b> Definición de las variables. ....	35
<b>Tabla 4.</b> Cronograma. ....	36
<b>Tabla 5.</b> Presupuesto. ....	37
<b>Tabla 6.</b> Valores de concordancia Kappa para todos los participantes al ser comparados con el Gold Standard (Patólogo con segunda especialidad en uropatología). ....	44
<b>Tabla 7.</b> Categorías asignadas a cada participante en las 2 fases del estudio según los valores Kappa, tal como fue definido por Landis JR et. al. (51). ....	46
<b>Tabla 8.</b> Distribución de los valores de Gleason por participante en la fase 1. ....	47
<b>Tabla 9.</b> Distribución de los valores de Gleason por participante en la fase 2. ....	48

**Lista de Apéndices**

**Apéndice A.** Consentimiento informado aplicado a todos los participantes del presente estudio.  
..... 62

**Apéndice B.** Modelo de la invitación para la convocatoria a la participación del estudio. .... 68

## Resumen

**Título:** Desempeño Diagnóstico y Reproducibilidad Interobservador del Grado de Gleason en la Interpretación Histopatológica Convencional y Mediante el Análisis de Imágenes por Software<sup>1</sup>

**Autor:** Carlos Andrés Sánchez Rueda<sup>2</sup>

**Palabras Clave:** Patología Digital, Histopatología, Gleason, Adenocarcinoma, Próstata, *Deep Learning*, Inteligencia Artificial.

### Descripción:

El adenocarcinoma de próstata produce gran morbimortalidad en la población mundial. El diagnóstico se basa en el grado Gleason definido por el patólogo. En el estudio histopatológico se reconocen múltiples fuentes de errores. El presente estudio evalúa el rendimiento diagnóstico y la concordancia interobservador de adenocarcinoma de próstata, y desarrolla una herramienta diagnóstica de inteligencia artificial para realizar predicciones en los grados Gleason de imágenes de patología digital para comparar el rendimiento diagnóstico y concordancia al hacer uso de ésta. Es un estudio observacional analítico prospectivo de concordancia, aplicado en un grupo de profesionales (patólogos generales, patólogo con segunda especialidad en patología y residente de patología), quienes realizaron marcaciones de grado Gleason en imágenes usando el *software QuPath*. Las marcaciones se analizaron obteniendo coeficientes de concordancia Kappa. Con estos datos, se desarrolla una herramienta de apoyo diagnóstico para la realización de predicciones usando inteligencia artificial sobre un último grupo de imágenes. Se evaluaron nuevamente coeficientes de concordancia Kappa para los participantes al definir grado Gleason conociendo la predicción del *software*. Los valores de Kappa para la primera fase del estudio oscilaron entre 0,276 y 0,537. Para la segunda fase, los valores Kappa oscilan entre 0,265 y 0,510, con mayor mejoría de rendimiento para el residente de patología en los valores Kappa (0,356 y 0,410), la sensibilidad y el valor predictivo positivo. Los coeficientes Kappa mostraron una discreta disminución para los patólogos. Se obtuvo un ascenso en la categoría de la interpretación de Kappa para 2 de los participantes de aceptable a moderado, con una mejoría mayor en el residente de patología. No se observó disminución en las categorías iniciales. Las herramientas de análisis de imágenes pueden ser útiles para entrenamiento de profesionales en patología.

---

<sup>1</sup> Trabajo de Grado.

<sup>2</sup> Facultad de Salud. Escuela de Medicina. Departamento de Patología. Directora Dra. Olga Mercedes Álvarez Ojeda. Codirectora Dra. Tania Mendoza Herrera.

## Abstract

**Title:** Diagnostic performance and interobserver reproducibility of Gleason grade in conventional histopathological interpretation and through software image analysis<sup>1</sup>

**Author:** Carlos Andrés Sánchez Rueda<sup>2</sup>

**Keywords:** Digital Pathology, Histopathology, Gleason, Adenocarcinoma, Prostate, Deep Learning, Artificial Intelligence.

### Description:

Prostate adenocarcinoma causes high morbidity and mortality in the world population. The diagnosis is based on the Gleason grade score defined by the pathologist. Multiple sources of errors have been found in the histopathological study process. The present study evaluates the diagnostic performance and interobserver agreement of prostate adenocarcinoma diagnosis and develops an artificial intelligence tool to make predictions on Gleason grades of digital pathology images to compare the diagnostic performance and agreement when using it. It is a prospective analytical observational concordance study, applied in a group of professionals (general pathologists, pathologist with a second specialty in uropathology, and pathology resident), who performed Gleason grade markings on images using QuPath software. The markings were analyzed obtaining Kappa concordance coefficients. With these data, a diagnostic support tool was developed to make predictions using artificial intelligence on a last group of images. The Kappa concordance coefficients for the participants were re-evaluated when defining the Gleason grade knowing the prediction of the software. Kappa values for the first phase of the study ranged from 0.276 to 0.537. For the second phase, the Kappa values ranged from 0.265 to 0.510, with greater improvement in performance for the pathology resident in the Kappa values (0.356 and 0.410), sensitivity, and positive predictive value. The Kappa coefficients showed a slight decrease for the pathologists. An upgrade in the Kappa interpretation category was obtained for 2 of the participants from fair to moderate, with a greater improvement in the pathology resident. No category decrease was observed. Image analysis tools could be useful for training pathology professionals.

---

<sup>1</sup> Bachelor Thesis.

<sup>2</sup> Faculty of Health. Medicine School. Department of Pathology. Director Dr. Olga Mercedes Álvarez Ojeda. Co-director Dr. Tania Mendoza Herrera.

### Introducción

El adenocarcinoma de próstata es un tipo de neoplasia que produce gran morbimortalidad en la población mundial donde se estiman 1.600.000 nuevos casos y 366.000 muertes anualmente. En nuestro país, se trata de la neoplasia en hombres que produce la segunda mortalidad más alta y que cuenta con la mayor incidencia (1), superando en número a los cánceres de pulmón, estómago y colorrectal.

El diagnóstico de cáncer de próstata se basa en el estudio histopatológico de la biopsia prostática y el hallazgo de células y glándulas neoplásicas en la muestra evaluada. Esta patología no puede ser diagnosticada con base en otros parámetros como lo son los niveles de Antígeno Prostático Específico (PSA, por sus siglas en inglés), examen físico, exámenes de laboratorio o sus síntomas (2). La escala de Gleason es el sistema más comúnmente usado para estadificar el adenocarcinoma de próstata (3), el cual se puede evaluar en material de biopsia mediante análisis histopatológico, obtenido a través de la sumatoria de los dos patrones predominantes en la muestra evaluada, oscilando entre 6 y 10 (4).

En el estudio histopatológico de cualquier pieza quirúrgica, se reconocen numerosas fuentes de posibles errores que pueden afectar la verosimilitud del diagnóstico emitido por el profesional en patología, dentro de estos, se encuentran problemas durante la fase analítica del diagnóstico (5), dados por sesgos de carácter cognitivo y visual (6). Los softwares de análisis de imágenes de patología digital como ayuda diagnóstica, han sido usados en múltiples estudios con diversas muestras de patología y piezas quirúrgicas, con resultados satisfactorios, debido a la facilidad en

su aplicación, su versatilidad y eficiencia (7-34). Esta herramienta representa un recurso que ofrece mayor objetividad al momento de evaluar la histopatología, limita el error humano y convierte al estudio histopatológico en un proceso replicable y medible.

El presente estudio pretende evaluar el rendimiento diagnóstico y la concordancia interobservador con el análisis digital de imágenes frente a la evaluación de imágenes histopatológicas de adenocarcinoma de próstata, mediante la aplicación de un algoritmo previamente entrenado que sugiere demarcaciones según la escala de Gleason realizado por profesionales en patología, patólogos con segunda especialización en uropatología y médicos residentes cursando el posgrado en esta especialidad médica.

### **1. Justificación**

El adenocarcinoma de próstata es una entidad de gran importancia por su alta morbimortalidad e incidencia, que la posicionan como la segunda neoplasia con mayor incidencia en el mundo y la segunda en mortalidad y primera en incidencia en hombres en nuestro país (1), con una incidencia cruda de 52,3 casos por cada 100.000 habitantes.

El diagnóstico de esta entidad requiere de una biopsia y el análisis histopatológico de un profesional en patología (médico patólogo o, idealmente, médico patólogo con segunda especialidad en uropatología) para su óptima identificación, durante lo cual se deberá clasificar la neoplasia mediante el uso de la escala de Gleason, que le confiere valor pronóstico al diagnóstico

emitido (4). En ocasiones, hay discrepancias en la estadificación y clasificación del estudio histopatológico entre los profesionales en patología, en función de la subjetividad que caracteriza a este proceso (5), lo que implica variaciones en el pronóstico y manejo de los pacientes afectados, por lo que es de vital importancia que éste sea lo más acertado posible.

La reciente introducción de métodos para el escaneo de láminas histopatológicas, ha permitido que la muestra antes vista de manera exclusiva mediante el microscopio, pueda ser evaluada en una imagen digitalizada de ésta (27–33), lo cual le permite al profesional en patología realizar su labor sin tener la lámina en su poder, otorgando el potencial de mejorar la oportunidad en el diagnóstico al no requerir el envío de láminas o, en su defecto, el desplazamiento del profesional al laboratorio o a la institución de salud donde se desempeña. En este sentido, la emergencia en salud producto de la pandemia declarada por la OMS en el año 2020 por cuenta de la infección por el SARS-CoV-2, supone una barrera importante para el desplazamiento, que representa un riesgo de contagio y posible exposición al agente infeccioso. Por tal motivo, reforzar recursos que involucren técnicas de patología digital se convierte en una necesidad imperiosa de nuestros tiempos.

Los software de análisis de imágenes, son recursos informáticos que tienen la posibilidad de realizar y orientar diagnósticos sobre imágenes histopatológicas digitalizadas mediante el reconocimiento e identificación de estructuras microscópicas de manera autónoma. Dada la evidencia disponible respecto al uso de métodos digitales de análisis de imágenes en patología y su capacidad para identificar patrones histopatológicos con el fin de facilitar el diagnóstico de diversas entidades (7–35), el presente estudio de investigación propone evaluar el uso de ambos

métodos (el tradicional, llevado a cabo por el patólogo mediante microscopía óptica, y el análisis digital de imágenes) de manera simultánea, entendiendo el segundo como un apoyo para el patólogo, con el fin de facilitar la labor del profesional y, en última instancia, identificar si se trata de una herramienta valiosa para dicho proceso.

## 2. Objetivos

### 2.1. Objetivo General

Comparar la capacidad discriminativa y la variabilidad interobservador de los grados de Gleason por medio de la interpretación histopatológica convencional versus el apoyado por software de análisis de imágenes (que desarrollará el Grupo de Investigación *Biomedical Imaging, Vision and Learning Laboratory* de la Escuela de Ingeniería de Sistemas de la UIS) (43), teniendo como *Gold Standard* el dictamen de un(os) profesional(es) experto(s) con segunda especialidad en uropatología, utilizando un banco de imágenes de libre uso (Dataset de imágenes disponible en la plataforma web de la Universidad de Harvard) (44).

### 2.2. Objetivos Específicos

- Describir la evaluación diagnóstica del *Gold Standard* o *Ground Truth*, dado por el médico patólogo con segunda especialidad en uropatología.
- Describir la participación del nivel de formación académica de los colaboradores en el rendimiento diagnóstico.
- Establecer el grado de concordancia interobservador entre la estadificación en la escala de antes de la intervención del software de reconocimiento de imágenes en el proceso diagnóstico histopatológico, discriminado por niveles de formación académica de los participantes (médicos residentes en patología, médicos patólogos generales o con estudios en áreas afines a la oncología urológica, y médicos con segunda especialidad en uropatología) y lo evaluado posterior a la sugerencia del software.

### **3. Marco Teórico**

#### **3.1. La Próstata**

La próstata es un órgano glandular que hace parte del aparato reproductor masculino. Es la glándula genital accesoria de mayor volumen y tiene como función producir un líquido alcalino rico en zinc y productos proteicos, con el objetivo de aportar nutrientes al espermatozoide y neutralizar el pH vaginal ácido. Se encuentra ubicada delante del recto, en posición caudal a la vejiga urinaria en la salida del tracto urinario rodeando la uretra. En dirección cefálica a la glándula prostática se encuentran las vesículas seminales, las cuales aportan la mayor parte del volumen del líquido seminal. La próstata está histológicamente constituida por 30 a 50 glándulas tubuloalveolares que vacían su contenido a la uretra prostática (45).

#### **3.2. El Adenocarcinoma de Próstata**

El adenocarcinoma de próstata es una enfermedad neoplásica que se presenta en hombres mayores de 50 años de edad, y es la forma de cáncer más frecuente en hombres en el mundo. En nuestro país, se trata de la neoplasia en hombres que produce la segunda mortalidad más alta y que cuenta con la mayor incidencia (1), superando en número a los cánceres de pulmón, estómago y colorrectal. A nivel global se estiman 1.600.000 nuevos casos y 366.000 muertes anualmente.

El adenocarcinoma de próstata tiene su origen en las glándulas presentes en dicho órgano. En el 70 – 80% de los casos, las glándulas neoplásicas tienen una posición periférica dentro de la próstata y, por tanto, síntomas como los obstructivos urinarios son tardíos, debido a lo cual la

identificación de esta neoplasia requiere de estrategias de cribado poblacional puesto que la captación con base en los síntomas dará como resultado la detección de casos con cáncer avanzado con un pobre pronóstico(46). El valor sérico de Antígeno Prostático Específico (PSA.), es la prueba de tamizaje más ampliamente usada para orientar la sospecha diagnóstica en pacientes masculinos en edad de riesgo. Sin embargo, esta prueba tiene limitaciones ampliamente conocidas, entre otras, que sus concentraciones se encuentran elevadas en la HBP (hiperplasia prostática benigna), la prostatitis, los infartos prostáticos, la instrumentación de la próstata y la eyaculación (46). Así mismo, entre un 20 a 40% de los pacientes con adenocarcinoma prostático no invasivo, presentan unos niveles de PSA en rango de normalidad. Aunque se han propuesto formas de mejorar la sensibilidad y especificidad de dicha prueba de tamizaje (estandarización por edad, relación PSA libre – PSA total, variabilidad en el tiempo de los valores), el PSA no permite realizar un diagnóstico definitivo y confiable, sino que debe ser interpretado como un paso previo que puede motivar la realización de una biopsia transrectal para estudio histopatológico (46).

Al evaluar histopatológicamente una biopsia de la próstata con el objetivo de definir la presencia de neoplasia, no sólo se debe confirmar o descartar la existencia de ésta, sino que es necesario graduar las características presentes en dicha neoplasia para orientar al profesional clínico en términos de pronóstico y gravedad.

### **3.3. Escala de Gleason**

La puntuación de Gleason es el sistema usado para graduar el adenocarcinoma de próstata. Es el resultado de la suma de los dos patrones de crecimiento tumoral más frecuentemente

evidenciados en la evaluación histopatológica (grados 1-5). Por lo tanto, el valor en la escala de Gleason para una muestra oscilará entre 2 y 10, siendo 2 el menos agresivo y 10 el más agresivo y de peor pronóstico (3)(47). Sin embargo, dado que se ha desincentivado el uso de las clasificaciones 1 y 2 para el score de Gleason por su tendencia a ser confundidos con lesiones benignas, se asume que el valor mínimo para el diagnóstico en una muestra será de 6 (3 + 3).

### ***3.3.1 Gleason Patrones 1 y 2***

Son patrones representados por nódulos bien circunscritos con glándulas pequeñas, bien delimitadas, uniformes, dispuestas en patrón “espalda con espalda”, sin infiltración evidente (grado 1) o presencia de infiltración focal alrededor de glándulas benignas (grado 2). Estos dos patrones son frecuentemente confundidos con otras lesiones benignas prostáticas puesto que sus características resultan sutiles. Recientemente, se ha acordado que estos patrones de la escala no deberían informarse en el estudio histopatológico debido a la dificultad que representa su correcta definición (4).

### ***3.3.2 Gleason Patrón 3***

Patrón compuesto exclusivamente por glándulas individuales bien formadas que demuestren un patrón de crecimiento infiltrativo entre glándulas benignas. Se trata de glándulas relativamente pequeñas, pero glándulas grandes también pueden estar presentes. Pueden existir escasas glándulas aparentemente fusionadas o pobremente formadas, lo cual es interpretado como un artefacto del seccionamiento tangencial de la muestra (4).

### **3.3.3 Gleason Patrón 4**

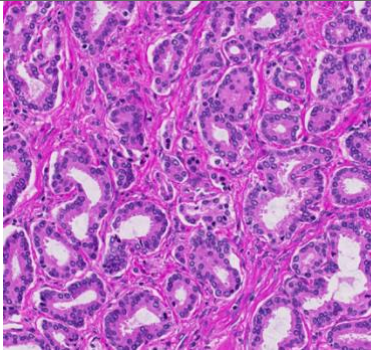
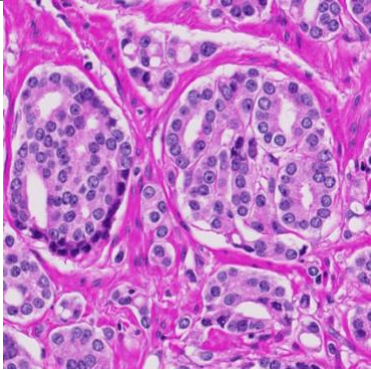
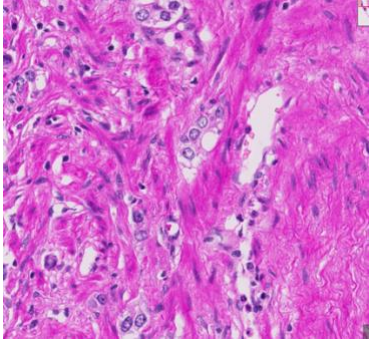
Contiene la mayor variabilidad en morfología con 4 patrones morfológicos principales: cribiforme, glomeruloide, fusionado, y glándulas pobremente formadas. Este diagnóstico debería ser visible a bajas magnificaciones (10X). Las glándulas glomeruloides consisten en pequeñas glándulas circulares malignas agrupadas dentro de una glándula dilatada y focalmente unidas a ella. El patrón cribiforme hace referencia a la presencia de glándulas con células epiteliales glandulares que ocupan su región central, dejando múltiples espacios o lúmenes separados dentro de ésta. Las glándulas pobremente formadas deberán verse agrupadas en *clusters* de al menos 10 de ellas que no estén inmediatamente adyacentes a otras glándulas de apariencia benigna (4).

### **3.3.4 Gleason Patrón 5**

Patrón de mayor agresividad. No muestra formación glandular. Se observarán células malignas que se agrupan formando sábanas o cordones que muestran infiltración. En algunas ocasiones es posible observar células en “anillo de sello” que podrían confundirse con el patrón cribiforme Gleason 4. La presencia de comedonecrosis es considerada criterio suficiente para clasificación Gleason 5, incluso si la misma se encuentra presente en glándulas cribiformes o glomeruloides (4).

## DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

**Tabla 1.** Resumen de los criterios usados para la clasificación mediante el score de Gleason para grados 3, 4 y 5, y ejemplos de imágenes histopatológicas extraídas del Dataset.

Patrón de Gleason	Descripción	Imagen
Gleason 3	Glándulas pequeñas e individuales no fusionadas y de patrón infiltrativo.	
Gleason 4	Contiene la mayor variabilidad en morfología con 4 patrones morfológicos principales: cribriforme, glomeruloide, fusionado, y glándulas pobremente formadas.	
Gleason 5	Células malignas que se agrupan formando sábanas o cordones que muestran infiltración. En algunas ocasiones es posible observar células en “anillo de sello”. Necrosis y comedonecrosis pueden ser hallazgos asociados.	

### 3.4. Factores de Riesgo

Los factores de riesgo que determinan la aparición de adenocarcinoma de próstata no han sido identificados con certeza. Sin embargo, se conocen 3 factores de riesgo bien descritos: edad

avanzada, origen étnico y herencia. Si un familiar en primer grado presenta cáncer de próstata, el riesgo de presentarlo se duplica. Si dos o más familiares en primer grado lo presentan, el riesgo asciende a ser 5 a 11 veces mayor. La incidencia de cáncer de próstata es elevada en regiones como Norteamérica y Europa septentrional, y baja en el sureste de Asia. Así mismo, los hombres de raza negra y la población escandinava presentan un mayor riesgo de padecer esta neoplasia (46)(47).

### **3.5. Patología Digital**

Tradicionalmente, la evaluación de los tejidos derivados de muestras en el ejercicio de la histopatología, se ha realizado principalmente a través de microscopía óptica *bright field*, es decir, mediante el uso de un microscopio óptico convencional donde se observa una muestra de tejido fijada en una lámina, la cual es observada por el patólogo de manera manual mediante la magnificación dada por el microscopio.

En los últimos años, la invención y disponibilidad en el mercado de escáneres de láminas histopatológicas, ha permitido su conversión en imágenes digitales (48)(49) (*Whole Slide Imaging*, WSI) para su posterior evaluación por los profesionales en patología a través de un ordenador u otro recurso electrónico similar, además de la posibilidad de que dicha evaluación pueda realizarse a la distancia, esto es, que el patólogo no requiere tener la lámina en su poder para evaluarla mediante microscopía. En este sentido, el bajo número de profesionales con segunda especialidad en todas las ramas de la medicina en países en vías de desarrollo como el nuestro da cuenta de la imperiosa necesidad de que herramientas como la patología digital jueguen un papel central en el proceso diagnóstico, puesto que permite zanzar barreras geográficas para el envío de láminas o el

desplazamiento del profesional para la evaluación de los casos. Así mismo, y como se mencionó anteriormente en la justificación de este estudio, la emergencia en salud producto de la pandemia declarada por la OMS en el año 2020 por cuenta de la infección por el SARS-CoV-2, supone una barrera importante para el desplazamiento, el cual representa un riesgo de contagio y posible exposición al agente infeccioso. De esta manera, reforzar recursos que involucren técnicas de patología digital se convierte en una necesidad imperiosa de nuestros tiempos.

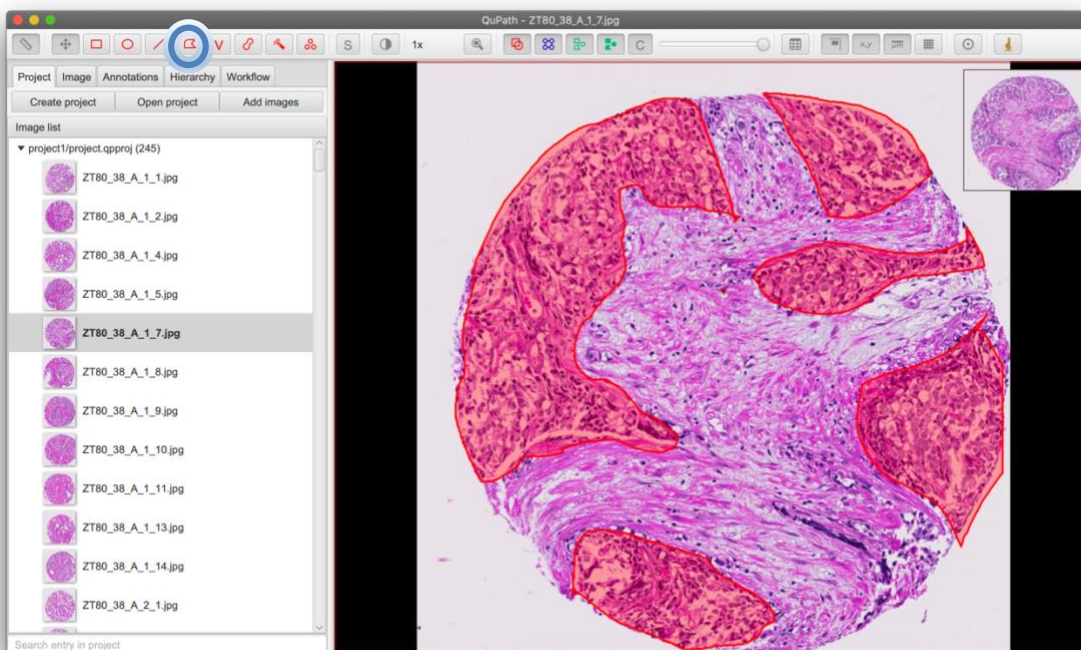
Múltiples estudios se han propuesto definir el rendimiento diagnóstico comparativo entre la evaluación microscópica tradicional y la realizada mediante la técnica de WSI (27–33) y la digitalización de otro tipo de imágenes como las de *tissue microarrays* (TMAs, microarreglos de tejido que muestran una imagen histopatológica de una región puntual de un tejido u órgano), y en la mayor parte de ellos los resultados han sido alentadores, encontrando una concordancia diagnóstica significativa entre dichos métodos. Estos estudios también destacan otra serie de ventajas que se obtienen al hacer uso de la patología digital, dentro de las cuales se encuentran: La posibilidad de realizar demarcaciones, etiquetado y resaltado de zonas de importancia y posteriormente compartirlas o fijarlas en la imagen; la facilidad para el almacenamiento de un archivo de patología en las instituciones y laboratorios; la durabilidad de dichas imágenes en el tiempo; y la facilidad para acceder a las mismas en caso de que se requiera una nueva evaluación de las mismas, entre otras.

### 3.6. Qupath y Demarcación de Imágenes de Patología Digital

QuPath es un programa informático que contiene, entre muchas otras funciones, la posibilidad de demarcar y etiquetar imágenes histopatológicas de WSIs o TMAs por parte de un evaluador, facilitando dicho proceso y permitiendo el posterior intercambio de las demarcaciones mediante la exportación de un archivo que contiene un “proyecto”, el cual corresponde al set de imágenes con sus respectivas demarcaciones.

Las demarcaciones serán etiquetadas como Gleason patrón 3 (Color azul), patrón 4 (Color amarillo) y patrón 5 (Color rojo). En la figura 1 se muestra un ejemplo.

*Figura 1. Captura de la ventana de QuPath.*



*Nota. Se muestra un ejemplo de TMA con una demarcación para áreas etiquetadas como Gleason patrón 5 (Áreas en rojo). El círculo azul señala el botón en la barra de herramientas correspondiente a la herramienta de demarcación que se usará para este propósito. En el lado izquierdo de la ventana, se encuentran las imágenes correspondientes al Set 1.*

### 3.7. Errores y Sesgos de la Evaluación Histopatológica Tradicional

En el estudio histopatológico de cualquier pieza quirúrgica, se reconocen numerosas fuentes de posibles errores que pueden afectar la verosimilitud del diagnóstico emitido por el profesional en patología, dentro de estos, se encuentran problemas durante la fase analítica del diagnóstico (5), dados por sesgos de carácter cognitivo y visual (6).

#### 3.7.1 Sesgos Cognitivos

Los sesgos cognitivos descritos en la práctica de la patología son:

- Sesgo de confirmación: Tendencia a la búsqueda de información y hallazgos que soporten o confirmen una hipótesis favorecida de manera previa.
- “Drift” diagnóstico: Cambios sutiles en los hallazgos con el tiempo que pueden producir que el evaluador no los note debido a su gradualidad.
- Anclaje: O visión de túnel. Se produce al prestar atención exclusivamente a algunos hallazgos en una evaluación que debe considerar otros aspectos. Habitualmente el fenómeno ocurre cuando los primeros hallazgos producen un anclaje de la impresión diagnóstica y se resta importancia a otras anomalías que no concuerden.
- Satisfacción de búsqueda: Suspensión de la evaluación cuando se encuentran suficientes hallazgos indicativos de un diagnóstico en particular.
- Sesgo de contexto: Fenómeno que se presenta cuando una muestra hace parte de un grupo en el que algún diagnóstico es especialmente frecuente, por lo que se asume *a priori* que el diagnóstico es éste, perdiendo objetividad en la búsqueda de los hallazgos.

- Evitación de rangos extremos: Se reconoce que existe una tendencia a clasificar diagnósticos como intermedios en una escala con mayor frecuencia, a pesar de que los hallazgos sean compatibles con un extremo bajo o alto en la misma.
- Preferencia de números: Se ha comprobado en múltiples estudios sobre variables numéricas continuas discretas que, a la hora de definir un hallazgo en valores numéricos, los evaluadores tienden a preferir números finalizados en 0 o 5. Por ejemplo, al estimar el peso de un paciente o al reportar el valor de tensión arterial en una consulta médica.
- Falacia del jugador: Es la tendencia a asumir que un resultado previo en una evaluación independiente afecta el resultado de otra evaluación independiente. Por ejemplo, se reconoce que, en la evaluación histopatológica, se da una renuencia del evaluador para asignar una clasificación particular en una escala diagnóstica si de manera previa ha clasificado varias muestras seguidas de esta misma manera.

### 3.7.2 *Sesgos Visuales*

Los sesgos visuales descritos en la práctica de la patología son:

- Ilusión de tamaño: Fenómeno presente cuando se subestima y sobreestima el tamaño de un objeto al encontrarse en presencia de otros objetos puesto que se tiende a hacer la comparación con éstos y no a juzgarlos de manera independiente.
- Ilusión de Craik-O'Brien-Cornsweet: Generalización en la definición de la diferencia del tono de dos colores al sólo juzgar la interfase entre ambos.
- Distinción de gradientes de colores: La distinción de la diferencia en la tonalidad de 2 o más colores, se facilita cuando se encuentran próximos entre sí, pero es especialmente

difícil cuando se valoran de manera independiente. Así mismo, la gama de tonos y colores perceptibles y diferenciables es limitada. Una persona adulta percibe una gama de unos 10 millones de colores, cuando una cámara de tan solo 12-bits es capaz de percibir más de 68.000 millones de colores.

- Inhibición lateral: Ilusión óptica en la cual se altera la percepción de colores al encontrarse en la vecindad de otros con intensidades muy disímiles. Ocurre debido a la inhibición de la actividad que realizan unas neuronas sobre otras evitando así la dispersión de la excitación hacia toda la retina.
- Ceguera por falta de atención: Ocurre cuando no se perciben algunos hallazgos a pesar de encontrarse dentro del rango visual, debido a que se está atendiendo a la búsqueda de otra anomalía distinta.

### **3.8. Análisis Digital de Imágenes**

Los softwares de análisis de imágenes de patología digital como ayuda diagnóstica, han sido usados en múltiples estudios con diversas muestras de patología y piezas quirúrgicas, con resultados satisfactorios, debido a la facilidad en su aplicación, su versatilidad y eficiencia (7-34). Son recursos informáticos que tienen la posibilidad de realizar y orientar diagnósticos sobre imágenes histopatológicas digitalizadas mediante el reconocimiento e identificación de estructuras microscópicas de manera autónoma. Consisten en la aplicación de un algoritmo o código en lenguaje de programación informática que puede aprender mediante el reconocimiento de patrones en las imágenes de patología digital, lo cual se logra al introducir información de demarcaciones o etiquetas que son puestas por un patólogo entrenado. Esta herramienta representa un recurso que

ofrece mayor objetividad al momento de evaluar la histopatología, limita el error humano y convierte al estudio histopatológico en un proceso replicable y medible.

Un estudio llevado a cabo por *E. Arvaniti y cols.* en 2018 (28), usó las demarcaciones de 2 patólogos en TMAs en imágenes de patología digital y comparó la concordancia interobservador en 641 imágenes. Posteriormente, se usaron dichas demarcaciones para entrenar una herramienta software capaz de reconocer regiones de interés en los TMAs y asignar un grado en la escala de Gleason de manera automatizada. El rendimiento del software fue evaluado en 245 nuevas imágenes de TMAs. Dentro de los hallazgos descritos, se encontró que la concordancia interobservador calculada mediante el coeficiente kappa de Cohen para el Gleason 3+3 fue de tan solo 0.47, valor que refleja una concordancia moderada que puede ser interpretada como subóptima para una prueba diagnóstica. Así mismo, los autores concluyen que el software así entrenado, alcanzó un nivel de discriminación y experticia similar al de los evaluadores. No obstante, indican que el software presentaba algunas limitaciones en relación con la identificación de algunas regiones en los TMAs que no correspondían a hallazgos típicos o que presentaban artefactos. En este sentido, los autores sugieren la necesidad de emprender estudios similares con un número mayor de participantes para permitir un entrenamiento mejor dirigido y nutrido por un número mayor de demarcaciones, que permita que el software que se entrene mediante *Deep learning* tenga una mejor capacidad de discriminación.

En otro estudio realizado por *J. Hipp y cols.* en 2012 (49), se usaron 2 algoritmos informáticos para identificar epitelio glandular maligno en imágenes histopatológicas de WSI de tejido prostático y resaltarlos dentro de la imagen, para posteriormente ser evaluado por los

profesionales en patología. Se reportan niveles de sensibilidad del 90% y especificidad del 90% para la identificación de las estructuras malignas con significado diagnóstico al compararlo con los patólogos participantes en el estudio. De esta manera, los autores proponen la implementación de software de características similares como apoyo para el trabajo del patólogo en el análisis de este tipo de muestras, puesto que facilitan la selección de los hallazgos relevantes, lo que se traduce en una mayor eficiencia en la labor del patólogo.

Tomando la experiencia del estudio de *E. Arvaniti y cols.* (28), un estudio que involucre una participación de un mayor número de evaluadores puede disminuir el margen de error que los autores indican presentó el software entrenado. Así mismo y como allí también se destaca, existe un nivel de concordancia subóptimo para el Gleason 3 por parte de los profesionales patólogos, lo que justificaría el uso de herramientas software como apoyo diagnóstico para el profesional.

### **4. Diseño Metodológico**

#### **4.1 Diseño del Estudio**

El presente es un estudio observacional analítico prospectivo de concordancia, aplicado en un grupo de profesionales entre los cuales se cuentan médicos patólogos generales, médico patólogo con segunda especialidad en uropatología y residente de patología, quienes emitieron diagnóstico (mediante demarcación de las áreas utilizando la herramienta QuPath previamente descrita) y estadificación mediante escala de Gleason (3) sobre imágenes digitalizadas de láminas de histopatología de libre uso (Dataset de imágenes disponible en la plataforma web de la

## DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

Universidad de Harvard) (44), de especímenes de adenocarcinoma de próstata, los cuales consisten en imágenes de alta calidad, procesadas con tinción de hematoxilina y eosina, todas ellas mediante técnicas de procesamiento idénticas, lo que las hace comparables para ser utilizadas en investigación.

Todos los participantes diligenciaron y firmaron un consentimiento informado diseñado específicamente para el presente estudio, donde también se recabaron datos sobre su nivel académico y estudios cursados afines a las áreas de patología quirúrgica, oncológica y uropatología.

Para la primera fase del estudio, 691 imágenes de TMAs se dividieron en 4 sets o grupos, siendo etiquetados por los participantes en el programa QuPath durante 2 o 3 semanas según el número de imágenes en cada set (Tabla 2).

**Tabla 2.** Características de los sets de imágenes histopatológicas que se usarán en la primera fase del estudio y el tiempo que tardará su demarcación.

Número del set de imágenes	Número de imágenes del set	Tiempo para la demarcación
<b>Grupo de imágenes 1 (Set 1)</b>	227 imágenes	3 semanas
<b>Grupo de imágenes 2 (Set 2)</b>	176 imágenes	2 semanas
<b>Grupo de imágenes 3 (Set 3)</b>	105 imágenes	2 semanas
<b>Grupo de imágenes 4 (Set 4)</b>	133 imágenes	2 semanas

Una vez se contó con las demarcaciones de los 4 sets mencionados anteriormente, el software de análisis de imágenes desarrollado por el grupo *Biomedical Imaging, Vision and Learning Laboratory* (43) fue entrenado usando los datos obtenidos.

En la segunda fase del estudio y en un nuevo grupo de imágenes histopatológicas compuesto por 245 imágenes en total, los participantes etiquetaron las imágenes histopatológicas y en cada una de ellas se indicó una demarcación sugerida por el software de análisis de imágenes. El participante contó con la posibilidad de seguir la sugerencia del *software* o realizar una demarcación alternativa, según su criterio. Ambos patrones (primario y secundario) definidos por el participante, fueron sujetos a análisis para evaluar la concordancia con el uropatólogo

colaborador en el presente estudio (Definido como *Gold Standard* o *Ground Truth* para este trabajo).

Se tuvo en cuenta en dicho análisis el nivel académico de los participantes y se estableció la concordancia mediante coeficiente de correlación Kappa de Cohen, y valores de sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN).

### **4.2. Población**

La población en estudio estuvo constituida por profesionales entre los cuales se cuentan médico residente de patología de la Universidad Industrial de Santander, médicos patólogos (Incluyendo quienes cuenten con títulos afines a patología quirúrgica y oncológica) y médico patólogo con segunda especialidad en uropatología, quienes aceptaron la participación en el estudio. De cada una de las categorías de nivel educativo, se esperó reclutar al menos un participante, sin que se limitara el volumen a un número máximo.

La participación de los profesionales en el estudio fue completamente voluntaria y cada participante contó con la posibilidad de desistir o retirarse en cualquier momento del mismo, sin que esto representara retaliaciones o repercusiones negativas.

#### **4.2.1 Criterios de Inclusión**

- Médicos patólogos que acepten la participación en el estudio.
- Médicos uropatólogos que acepten la participación en el estudio.

- Médicos residentes de la Universidad Industrial de Santander de primero, segundo y tercer año que acepten la participación en el estudio.

### 4.2.2 *Criterios de Exclusión*

- Una vez se realizó el entrenamiento en la herramienta de demarcación QuPath a los participantes, se estableció como criterio de exclusión imposibilidad para usar dicho recurso por limitaciones técnicas o de otro tipo.

### 4.3. Instrumentos del Estudio

Para la realización del presente estudio se usaron imágenes digitalizadas de TMAs que histopatológicamente corresponden a lesiones sugestivas de adenocarcinoma de próstata del dataset de imágenes disponible en la plataforma web de la Universidad de Harvard (44), las cuales fueron evaluadas por los participantes para definición de clasificación y estadio de la neoplasia. Se hizo uso del software *QuPath* (50), herramienta de procesamiento y etiquetado de imágenes histopatológicas gratuita y de libre uso, para la demarcación de los TMAs por parte de los participantes.

En la primera fase, los participantes realizaron el proceso de demarcación para los TMAs que se usaron en el estudio. Posteriormente, y en un nuevo grupo de imágenes (segunda fase del estudio), se utilizó el software de análisis de imágenes desarrollado por el grupo de investigación *Biomedical Imaging, Vision and Learning Laboratory* (43), grupo de bioingeniería de la Universidad Industrial de Santander, el cual indicó una demarcación alternativa sugerida al

profesional, con base en el entrenamiento obtenido a través de los diagnósticos emitidos en las demarcaciones dadas por los especialistas. Se hizo uso de este software de análisis de imágenes previo acuerdo con el grupo de investigación de ingeniería biomédica.

### **4.4. Procesamiento de Datos**

La información obtenida de todos los instrumentos mencionados, extraída de los mismos por los integrantes del grupo de investigación de bioingeniería, se digitó en una hoja electrónica Excel, registrando los valores para todas las variables en estudio. Posteriormente se exportaron y analizaron los datos con STATA 12.0 para el análisis de correlación y concordancia, según lo que se describe en el apartado “Análisis estadístico de las variables”.

## **5. Evaluación Crítica del Protocolo**

### **5.1. Ventajas del Diseño**

- Las imágenes histopatológicas que se usaron son comparables entre sí y cuentan con técnicas de procesamiento similares que permiten usarlas como punto de referencia para el estudio.
- El programa informático QuPath, utilizado para la demarcación de las imágenes histopatológicas, se encuentra especialmente diseñado para el uso en recursos de patología digital, facilitando así su utilidad y manejo.
- El análisis comparativo de las imágenes se realizó mediante un software informático que midió las áreas demarcadas para cada patrón de Gleason en cada una de las imágenes

evaluadas, lo que permitió una mayor objetividad en los patrones definidos como primario y secundario.

- Todos los diferentes pasos del presente estudio se realizaron mediante recursos digitales, lo que facilitó el intercambio de los datos y las demarcaciones, evitando así el desplazamiento y contacto físico entre participantes; o el intercambio de objetos en el contexto de la pandemia por SARS-CoV-2, durante la cual tuvo lugar este estudio.

### **5.2. Desventajas del Diseño**

- Las demarcaciones realizadas por los participantes aún pueden presentar errores humanos al momento de trazar los límites entre las diferentes zonas definidas para el score de Gleason, lo cual podría inducir al error en el posterior análisis de datos.
- Debido a que se requiere la evaluación de una cantidad significativa de imágenes histopatológicas, se corría el riesgo de producirse una pérdida de participantes que no estuvieran dispuestos a continuar con el estudio debido al tiempo que exige la demarcación requerida.
- Los profesionales podrían haber encontrado dificultades al usar los programas informáticos descritos si no se contaban con habilidades mínimas requeridas para manejar un ordenador.

## **6. Análisis Estadístico de las Variables**

Se utilizaron coeficientes Kappa de Cohen para evaluar la variabilidad de la puntuación por score de Gleason entre observadores. El coeficiente Kappa de Cohen es una medida de

concordancia que intenta estimar el nivel de acuerdo entre dos observadores excluyendo la probabilidad del acuerdo aleatorio y presenta valores entre 0 y 1; se acerca a 1 cuanto mayor nivel de acuerdo entre 2 observadores. Los valores Kappa midieron la concordancia entre observadores para ambas fases del estudio. Utilizando todas las lesiones identificadas por los lectores, se estableció la sensibilidad, especificidad, VPP y VPN con intervalo de confianza (IC) del 95% para cada nivel de puntuación en cada lector del software de análisis de imágenes.

### 7. Definición de las Variables

*Tabla 3. Definición de las variables.*

Nombre de la variable	Definición Conceptual	Definición Operacional	Escala de medición
<b>Nivel académico del participante.</b>	Último nivel académico alcanzado por el participante incluyendo estudios en áreas afines a la patología quirúrgica, oncológica y uropatología.	Definido como médico residente de patología, médico patólogo (General, con estudios en patología quirúrgica o patología oncológica) o médico patólogo con segunda especialidad en uropatología.	Variable cualitativa nominal.
<b>Grado de Gleason</b>	Definido como el grado histopatológico otorgado a un área demarcada en la imagen histopatológica.	Puede asumir los valores: Gleason 3, Gleason 4 o Gleason 5.	Variable cualitativa ordinal.

## 8. Cronograma

*Tabla 4. Cronograma.*

Actividades	Julio a Agosto 2020	Septiembre a Octubre 2020	Noviembre a Diciembre 2020	Enero a Junio 2021	Julio a Octubre 2021	Noviembre a Diciembre 2021	Enero a Agosto 2022	Septiembre a Octubre 2022	Noviembre a Diciembre 2022
Elaboración protocolo									
Sustentación y presentación al CEINCI									
Artículo de revisión									
Recolección de datos									
Tabulación de datos									
Análisis de datos									
Creación documento de resultados									
Publicación de resultados									
Presentación en eventos y redacción de artículo para revista									

## 9. Presupuesto

*Tabla 5. Presupuesto.*

<b>Concepto</b>	<b>Valor pasivo</b>
Servicios prestados por la directora	\$8.960.000 anual
Servicios prestados por la codirectora	\$4.480.000 anual
Gastos por impresiones	\$50.000
Computador	\$350.000
Compra de equipos	-
Alquiler de equipos	-
Licencia de software	-
Bibliografía	\$3.000.000
Publicación en revistas	\$2.000.000
Divulgación y socialización de resultados	\$1.000.000
Participación en eventos académicos (Nacionales e internacionales)	\$5.000.000
<b>Total</b>	<b>\$24.840.000</b>

## 10. Consideraciones Éticas

### 10.1. Principios Éticos

A continuación, se describen los principios éticos.

#### 10.1.1 Principio de Beneficencia

El presente proyecto de investigación pretendió recolectar evidencia respecto a la utilidad de una herramienta en el ámbito de la patología digital lo cual tiene el potencial de mejorar la oportunidad en el diagnóstico de adenocarcinoma de próstata y el potencial entrenamiento de nuevos profesionales en patología y, mediante el uso de un software de análisis de imágenes, establecer si representa un recurso que permita mejorar la eficiencia, reproducibilidad y objetividad de la evaluación histopatológica.

En la muestra se utilizaron imágenes histopatológicas de próstata, de pacientes cuya identidad es desconocida para el autor y todos los participantes en el estudio, por lo cual los datos obtenidos no generan perjuicios potenciales para dichos pacientes. Así mismo, y como se ha anotado en el presente protocolo, las imágenes son de dominio público, lo cual se encuentra establecido en la plataforma web de la universidad de Harvard (44). Dichas imágenes histopatológicas provienen originalmente del Hospital Universitario de Zurich, Suiza.

### *10.1.2 Principio de Respeto a las Personas*

La población de profesionales que participaron en el estudio firmaron un consentimiento informado donde aceptaron dicha participación y mediante el cual se les informaron consideraciones respecto al manejo de los datos obtenidos.

Todos los participantes contaron con la protección de sus identidades, puesto que el autor del estudio, el asesor epidemiológico y el director del proyecto de grado desconocen dichos datos. Los participantes fueron identificados con un código asignado por parte del grupo de bioingeniería al recibir las demarcaciones provenientes de cada uno de dichos profesionales y, posteriormente, enviaron los datos obtenidos al autor y asesor epidemiológico para su respectivo análisis estadístico.

### *10.1.3 Principio de Justicia*

La única distinción considerada en el presente estudio es el nivel académico de los participantes para el proceso de determinar asociaciones de rendimiento diagnóstico con dicha variable. Sin embargo, la participación no se encontró condicionada por ninguna distinción de raza, género, sexo, edad, etnia o clase social. En este sentido, no se incumple el principio de justicia pues se dieron las mismas garantías de protección de la identidad y de los datos a todos los participantes y se consolidará la voluntad de participar mediante un consentimiento informado para la totalidad de la población.

### 10.2. Tratamiento de los Datos Personales

La confidencialidad de los datos de los profesionales médicos residentes, especialistas en patología y médicos patólogos con segunda especialidad, se garantizó mediante la asignación de un código a cada participante, lo que constituye un ciego para todos los involucrados en la presente investigación incluyendo al autor del estudio, el asesor epidemiológico y el director del presente proyecto. La identidad de los participantes sólo fue conocida por profesionales del grupo de bioingeniería quienes establecieron el código que se usó en adelante.

No se incluyeron variables de carácter personal como sexo, género, edad, etnia o clase social que pudieran generar discriminación. La única variable que se tuvo en cuenta fue el nivel académico de los participantes para efectos de la determinación de asociaciones de rendimiento diagnóstico.

En concordancia con la Ley 1581 de 2012, ante solicitud expresa del participante será eliminada su identidad, y demás datos recabados, de la base de datos que se construyó a partir de este estudio, previo consentimiento informado de todos los participantes (Apéndice A). Dicha base de datos fue manejada por el grupo de ingeniería biomédica *Biomedical Imaging, Vision and Learning Laboratory* quienes asignaron los códigos a cada participante. El documento que la contiene se encuentra en un archivo dispuesto en un servidor del laboratorio del grupo de investigación de ingeniería biomédica, el cual está protegido por una clave de acceso que impide su visualización por personas no autorizadas. Una vez el estudio se dé por finalizado, los datos recabados (Nivel académico del participante y demarcaciones realizadas) sin incluir la identidad de los participantes, se dispondrá como datos de dominio público por parte del grupo de ingeniería,

donde podrán ser usados por diversos grupos de investigación afines, incluyendo los grupos de investigación de ingeniería biomédica y patología de la Universidad Industrial de Santander. Dicha disposición final de los datos pretende permitir la generación de otros proyectos similares que motiven la creación de herramientas en patología digital.

### **10.3. Clasificación del Riesgo de la Investigación**

Según lo establecido en el artículo 11 de la resolución 8430 de 1993 del Ministerio de Salud, el presente estudio representa una investigación SIN RIESGO, dado que no se realizó ninguna intervención o modificación intencionada de las variables biológicas, fisiológicas o psicológicas de los individuos que participan en el estudio.

### **10.4. Procedimientos para Abordar a los Participantes**

Se realizó una reunión convocada por el autor, director del proyecto y asesor epidemiológico, donde se invitó a diferentes médicos especialistas en patología y médico patólogo con segunda especialidad en uropatología, estableciendo las características del proyecto, quienes manifestaron su voluntad de participar, diligenciaron un consentimiento informado donde se consignó el deseo de hacer parte del proceso. En adelante, cada participante realizó las demarcaciones y las envió a los autores del estudio, quienes, a su vez, replicaron los datos al grupo de ingeniería biomédica. Los participantes tenían la posibilidad de interrumpir su participación en cualquier momento y fase del estudio sin que esto implicara consecuencias negativas. A través de la participación, se indicó la posibilidad de ser considerado para posteriores proyectos de la línea de investigación relacionada con el grupo de ingeniería biomédica de la universidad y el

departamento de patología. Un modelo de la invitación para la convocatoria se incluye en el Apéndice B.

### **10.5. Potenciales Beneficiarios**

El presente proyecto de investigación pretendió recolectar evidencia respecto a la utilidad de una herramienta en el ámbito de la patología digital lo cual tiene el potencial de mejorar la oportunidad en el diagnóstico de adenocarcinoma de próstata y el potencial entrenamiento de nuevos profesionales en patología y, mediante el uso de un software de análisis de imágenes, establecer si representa un recurso que permita mejorar la eficiencia, reproducibilidad y objetividad de la evaluación histopatológica. En este sentido, la población potencialmente beneficiaria sería todo paciente masculino en edad de riesgo de adenocarcinoma de próstata a quien se le tome una muestra para la realización de la evaluación histopatológica en el tejido obtenido.

Desde el punto de vista de los profesionales en patología, existe un beneficio potencial dado que, al facilitarse el reconocimiento de regiones de interés y contar con una orientación diagnóstica, se contaría con una mayor precisión y eficiencia del proceso diagnóstico, lo que optimizaría dicho proceso.

Así mismo, dado que el software ha recibido un entrenamiento previo a través de demarcaciones de profesionales expertos, puede representar un apoyo para el aprendizaje de médicos residentes en patología o quienes se encuentren cursando una segunda especialidad en uropatología o ramas afines, puesto que se trata de un recurso que tiene la capacidad de sugerir un diagnóstico en una

imagen histopatológica en tiempo real, permitiendo la realización del ejercicio diagnóstico por parte del estudiante para posteriormente recibir la retroalimentación del software a modo de orientación.

### **10.6. Riesgos para los Participantes e Investigadores**

Los riesgos para los participantes podrían encontrarse relacionados con el manejo de los datos recabados de cada uno de ellos, dentro de los cuales se cuentan exclusivamente la identificación (Nombre) y el nivel académico. En consecuencia, se estableció un ciego para los investigadores, de tal forma que se desconozca la identidad de los participantes en relación con los datos obtenidos. La única variable que se encontró a disposición de los investigadores fue la del nivel académico puesto que se relacionó con el rendimiento diagnóstico. No se consideraron riesgos para los investigadores en el presente estudio.

### **10.7. Consentimiento Informado**

El consentimiento informado que se usó en el presente estudio se encuentra en el apartado Apéndice A.

## **11. Resultados**

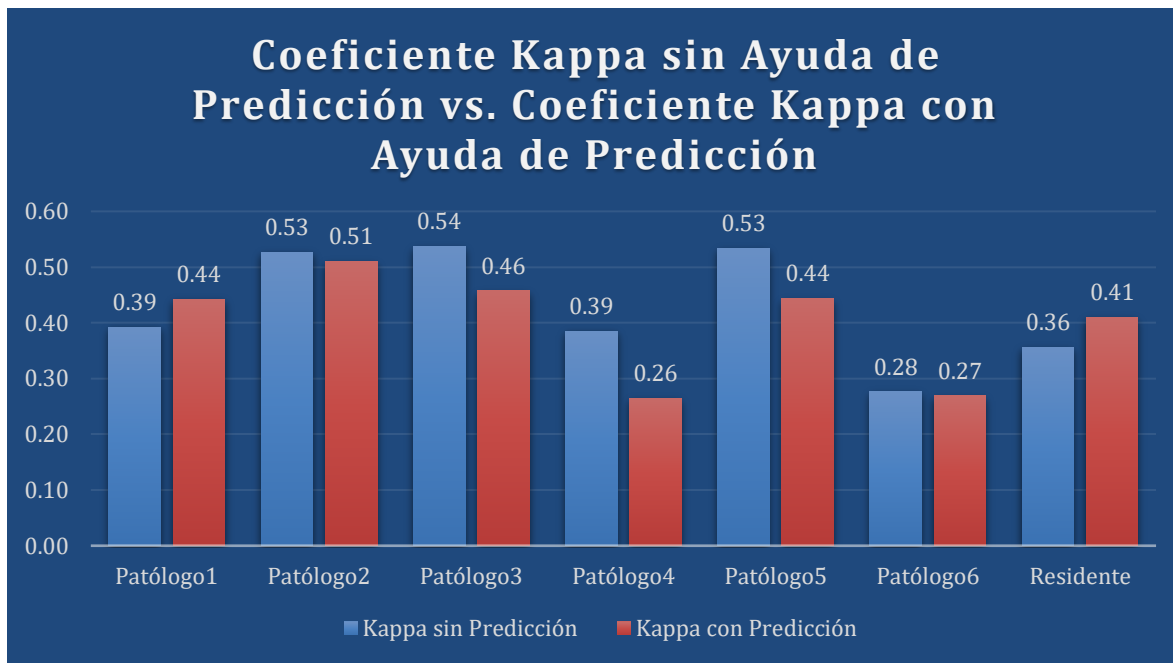
Los valores de los coeficientes Kappa para cada uno de los participantes en el estudio en los primeros 4 grupos de imágenes histopatológicas y en el último grupo de imágenes con ayuda de las predicciones se resumen en la tabla 6 y en la figura 2.

**Tabla 6.** Valores de concordancia Kappa para todos los participantes al ser comparados con el Gold Standard (Patólogo con segunda especialidad en uropatología).

<i>Participante</i>	<i>Kappa sin Predicción</i>	<i>Kappa con Predicción</i>	<i>Diferencia</i>
<i>Patólogo 1</i>	0,392	0,442	0,0500
<i>Patólogo 2</i>	0,527	0,510	-0,0174
<i>Patólogo 3</i>	0,537	0,457	-0,0794
<i>Patólogo 4</i>	0,385	0,265	-0,1204
<i>Patólogo 5</i>	0,534	0,443	-0,0906
<i>Patólogo 6</i>	0,276	0,269	-0,0076
<i>Residente</i>	0,356	0,410	0,0538

Los valores de Kappa para la primera fase del estudio oscilaron entre 0,276 y 0,537. Para la segunda fase, los valores Kappa oscilaron entre 0,265 y 0,510. Para la mayoría de los participantes, se obtuvo una ligera disminución del valor Kappa al contar con la predicción del software; en el caso del patólogo 1 y del residente, se reporta un aumento del coeficiente de Kappa para la segunda fase.

**Figura 2.** Gráfica de barras que compara el coeficiente Kappa sin ayuda de la predicción y el coeficiente Kappa con ayuda de la predicción generada mediante inteligencia artificial.



Tal como fue definido por Landis JR et. al. (51), el coeficiente Kappa puede dividirse en rangos para su interpretación, así: 0 pobre; 0,01-0,2 leve; 0,21-0,4 aceptable; 0,41-0,6 moderado; 0,61-0,8 considerable; 0,81-1 casi perfecta.

En este sentido, los resultados obtenidos en la primera fase del estudio pueden interpretarse como aceptables para 4 de los participantes (patólogo 1, patólogo 4, patólogo 6 y residente), son moderados para 3 de los participantes (patólogo 2, patólogo 3 y patólogo 5). En la segunda fase del estudio, los resultados pueden interpretarse como aceptables para 2 participantes (patólogo 4 y patólogo 6) y moderados para 5 participantes (patólogo 1, patólogo 2, patólogo 3, patólogo 5 y residente), tal como se expone en la tabla 7.

*Tabla 7. Categorías asignadas a cada participante en las 2 fases del estudio según los valores Kappa, tal como fue definido por Landis JR et. al. (51).*

<b>Participante</b>	<b>Kappa sin Predicción</b>	<b>Kappa con Predicción</b>
<b>Patólogo1</b>	Acceptable	Moderado
<b>Patólogo2</b>	Moderado	Moderado
<b>Patólogo3</b>	Moderado	Moderado
<b>Patólogo4</b>	Acceptable	Acceptable
<b>Patólogo5</b>	Moderado	Moderado
<b>Patólogo6</b>	Acceptable	Acceptable
<b>Residente</b>	Acceptable	Moderado

Se calcularon los valores para sensibilidad, especificidad, VPP y VPN para cada uno de los participantes según la fase del estudio. Los valores para las fases 1 y 2 se resumen en la tabla 8 y tabla 9 respectivamente.

## DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

**Tabla 8.** Distribución de los valores de Gleason por participante en la fase 1.

Distribución/Valores Concordantes	UP	P1	P2	P3	P4	P5	P6	Residente
Gleason 0+0	128	101	122	108	101	102	104	59
Gleason 3+3	192	72	111	98	76	141	68	70
Gleason 3+4	47	1	2	18	3	4	2	8
Gleason 3+5	2	0	0	0	0	0	0	0
Gleason 4+3	25	3	1	18	3	10	1	7
Gleason 4+4	97	73	50	49	27	68	19	46
Gleason 4+5	6	0	0	1	1	0	0	3
Gleason 5+3	1	0	0	0	0	0	0	0
Gleason 5+4	4	0	1	3	0	1	0	1
Gleason 5+5	91	52	84	52	82	43	56	71
Total Benigno concordante	128	101	122	108	101	102	104	59
Total Benigno	128	115	209	115	120	103	148	75
Total Maligno concordante	465	201	249	239	192	267	146	206
Total Maligno	465	478	384	478	473	490	445	518
Total Concordante	593	302	371	347	293	369	250	265
Sensibilidad		0,432	0,535	0,514	0,413	0,574	0,314	0,443
Especificidad		0,789	0,953	0,844	0,789	0,797	0,813	0,461
VPP		0,421	0,648	0,500	0,406	0,545	0,328	0,398
VPN		0,878	0,584	0,939	0,842	0,990	0,703	0,787

*Nota.* Se indican los valores para la sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN), para los participantes: Uropatólogo (UP), patólogos (P1, P2, P3, P4, P5, P6) y residente.

DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

*Tabla 9. Distribución de los valores de Gleason por participante en la fase 2.*

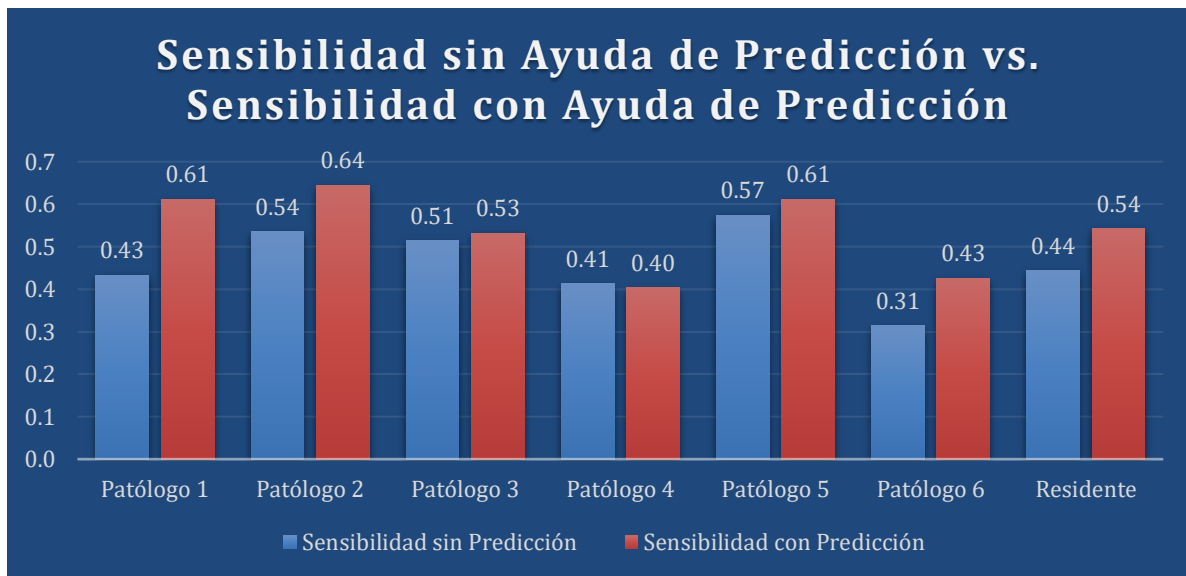
Distribución/Valores Concordantes	UP	P1	P2	P3	P4	P5	P6	Residente
Gleason 0+0	15	12	12	11	8	10	8	8
Gleason 3+3	70	46	44	44	35	38	32	26
Gleason 3+4	21	0	2	3	2	2	0	8
Gleason 3+5	0	0	0	0	0	0	0	0
Gleason 4+3	16	0	4	4	2	2	0	7
Gleason 4+4	90	80	78	53	34	80	45	66
Gleason 4+5	3	0	0	0	1	0	0	1
Gleason 5+3	0	0	0	0	0	0	0	0
Gleason 5+4	1	0	0	0	0	0	0	0
Gleason 5+5	15	6	11	11	13	10	15	9
Total Benigno concordante	15	12	12	11	8	10	8	8
Total Benigno	15	20	20	11	13	12	14	17
Total Maligno concordante	216	132	139	115	87	132	92	117
Total Maligno	216	211	211	220	218	219	217	214
Total Concordante	231	144	151	126	95	142	100	125
Sensibilidad		0,611	0,644	0,532	0,403	0,611	0,426	0,542
Especificidad		0,800	0,800	0,733	0,533	0,667	0,533	0,533
VPP		0,626	0,659	0,523	0,399	0,603	0,424	0,547
VPN		0,600	0,600	1,000	0,615	0,833	0,571	0,471

*Nota. Se indican los valores para la sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN), para los participantes: Uropatólogo (UP), patólogos (P1, P2, P3, P4, P5, P6) y residente.*

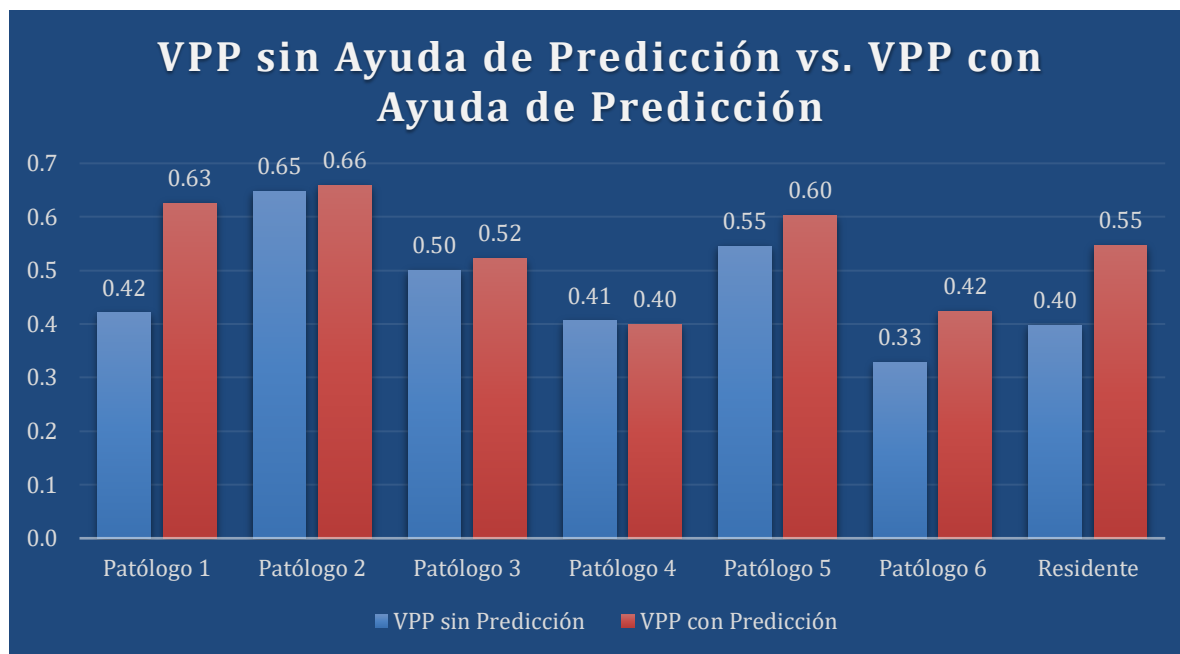
Es importante destacar que en la fase 2 el volumen de TMA's negativos para adenocarcinoma o "benignos" es limitada (15 TMA's en total, como puede verse en la casilla "Total

Benigno” para el uropatólogo en la tabla 9), razón por la cual los valores de especificidad y VPN deben interpretarse con cautela, puesto que carecen de potencia estadística. Por este motivo, se ilustran solamente los valores de sensibilidad y VPP en las figuras 3 y 4.

**Figura 3.** Gráfica de barras que compara la sensibilidad sin ayuda de la predicción y la sensibilidad con ayuda de la predicción generada mediante inteligencia artificial.



**Figura 4.** Gráfica de barras que compara el VPP sin ayuda de la predicción y el VPP con ayuda de la predicción generada mediante inteligencia artificial.



Como se puede observar en la gráfica representada en la figura 3, la sensibilidad obtuvo un ascenso en 6 de los 7 participantes. Para el residente, la sensibilidad aumentó 0.11, lo cual representa una mejoría de un 25% respecto al valor inicial. Para los médicos patólogos la sensibilidad aumentó en 5 de los 6 participantes para esta población, obteniendo un aumento promedio de 0.075 para la sensibilidad, lo cual representa una mejoría de un 16,24% respecto a la sensibilidad de la fase sin predicción del *software*.

Análogamente, se puede observar en la gráfica representada en la figura 4 que el VPP obtuvo un ascenso en 6 de los 7 participantes. Para el residente, el VPP aumentó 0.15, lo cual representa una mejoría de un 37.5% respecto al valor inicial. Para los médicos patólogos el VPP aumentó en 5 de los 6 participantes para esta población, obteniendo un aumento promedio de 0.062 para el

VPP, lo cual representa una mejoría de un 13% respecto al VPP de la fase sin predicción del *software*.

### 12. Discusión y Conclusión

Se obtuvo una discreta disminución en los coeficientes Kappa de Cohen para la mayoría de los participantes en este estudio al usar una herramienta informática de tipo *Software* de inteligencia artificial con predicciones diagnósticas en comparación con el diagnóstico emitido por los mismos participantes sin la ayuda de la herramienta. Sin embargo, en 2 de los participantes, se obtuvo una mejoría de dicho coeficiente que significó un ascenso en la categoría de la lectura de aceptable a moderada para los valores Kappa. El residente de patología obtuvo una mayor mejoría de rendimiento diagnóstico según el valor Kappa (0,356 en primera fase y 0,410 en segunda fase). El VPP y la sensibilidad presentan una mejoría en 6 de los 7 participantes como se ilustra en las imágenes 3 y 4.

En el estudio de Yelda Dere et. al., titulado *A grading dilemma; Gleason scoring system: Are we sufficiently compatible? A multi center study* (52), se compara la concordancia interobservador para el valor de Gleason asignado por 8 patólogos a láminas histopatológicas de próstata, obteniendo un valor Kappa de 0,44. En otro estudio de Tayyar A. Ozkan et. al., titulado *Interobserver variability in Gleason histological grading of prostate cancer* (53), se compara la concordancia interobservador de 2 patólogos para 407 biopsias de próstata, obteniendo un valor Kappa de 0,43. En el estudio de Veloso et. al., titulado *Interobserver agreement of Gleason score*

*and modified Gleason score in needle biopsy and in surgical specimen of prostate cancer* (54), comparan la concordancia interobservador de 3 patólogos para 110 casos de biopsia de próstata, obteniendo un valor Kappa de 0,35. Finalmente, en otro estudio de Allsbrook et al., titulado *Interobserver reproducibility of Gleason grading of prostatic carcinoma: General pathologists* (55), comparan la concordancia interobservador de 41 patólogos para 38 casos de biopsia de próstata, obteniendo un valor Kappa de 0,43.

Como se puede observar, la concordancia interobservador medida mediante valores Kappa, se mantiene en el rango de aceptable a moderado para los estudios que evalúan esta variable. En nuestro estudio, como se mencionó anteriormente, los valores de Kappa para la primera fase del estudio oscilaron entre 0,276 y 0,537 y para la segunda fase, los valores Kappa oscilaron entre 0,265 y 0,510, rangos dentro de los cuales se encuentran los valores usuales para el Kappa cuando se evalúa la concordancia interobservador en el *score* de Gleason en otros estudios.

Es necesario ampliar este tipo de trabajos con más participantes, y diseñar herramientas con mayor número de imágenes para mejorar la calidad de las predicciones en relación con el uso de ayudas diagnósticas para el patólogo que permitan mejorar la concordancia interobservador. En nuestro estudio, la interpretación de los resultados estuvo limitada por el bajo volumen de imágenes de TMAs benignos en la segunda fase, por lo cual, en nuevos diseños de estudios, puede ser importante contar con la lectura del patólogo experto de referencia o *Gold Standard* con anterioridad a la lectura de los participantes, para aleatorizar la distribución de las imágenes en los diferentes grupos o fases.

Se reconoce una limitación para el uso de herramientas de inteligencia artificial en nuestro medio, dado que no se cuenta con una estrategia de digitalización de imágenes en la mayoría de instituciones, puesto que no existen escáneres de láminas histopatológicas. El hecho de que se desarrollen este tipo de herramientas que permiten una mejoría en términos de eficiencia y aumento de rendimiento reflejado en variables como la sensibilidad y el VPP, puede ser un incentivo para la introducción de la digitalización de láminas histopatológicas. Así mismo, es posible, mediante herramientas automatizadas, establecer programas que aporten una lectura inicial por parte del *software* que permita, a su vez, la priorización de casos que supongan una mayor probabilidad de resultado maligno, en laboratorios o instituciones con alto volumen de casos.

Las herramientas de análisis de imágenes por software podrían ser útiles para el entrenamiento de profesionales en patología, dada la mejoría para los valores Kappa, la sensibilidad y el VPP obtenidos en el participante con menor experiencia en este estudio (residente de patología). Por lo tanto, dentro de las herramientas de simulación para enseñanza con métodos digitales en los estudiantes de posgrado en formación, puede ser relevante tener en cuenta estrategias de aprendizaje mediante *software* de inteligencia artificial.

### **13. Conflicto de Intereses**

Los autores de este estudio no manifiestan conflictos de intereses.

### Referencias Bibliográficas

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. Philip W Kantoff, MDMary-Ellen Taplin, MDJoseph A Smith M. Clinical presentation and diagnosis of prostate cancer. UpToDate [Internet]. 2020; Disponible en: <https://www.uptodate.com/contents/clinical-presentation-and-diagnosis-of-prostate-cancer>
3. Gleason DF, Mellinger GT, Arduino LJ, Bailar JC, Becker LE, Berman HI, et al. Prediction of Prognosis for Prostatic Adenocarcinoma by Combined Histological Grading and Clinical Staging. *J Urol.* 2017;197(2):S134–9.
4. Braunhut BL, Punnen S, Kryvenko ON. Updates on Grading and Staging of Prostate Cancer. *Surg Pathol Clin.* 2018;11(4):759–74.
5. Aeffner F, Wilson K, Martin NT, Black JC, Hendriks CLL, Bolon B, et al. The gold standard paradox in digital image analysis: Manual versus automated scoring as ground truth. *Arch Pathol Lab Med.* 2017;141(9):1267–75.
6. Sirota RL. Defining error in anatomic pathology. *Arch Pathol Lab Med.* 2006;130(5):604–6.
7. Ozerdem U, Wojcik EM, Barkan GA, Duan X, Erşahin Ç. A practical application of quantitative vascular image analysis in breast pathology. *Pathol Res Pract.* 2013;209(7):455–8.
8. Słodkowska J, Markiewicz T, Grala B, Kozłowski W, Papierz W, Pleskacz K, et al.

- Accuracy of a remote quantitative image analysis in the whole slide images. *Diagn Pathol.* 2011;6(SUPPL. 1):2–6.
9. Maeda I, Abe K, Koizumi H, Nakajima C, Tajima S, Aoki H, et al. Comparison between Ki67 labeling index determined using image analysis software with virtual slide system and that determined visually in breast cancer. *Breast Cancer.* 2016;23(5):745–51.
  10. Mohammed ZMA, McMillan DC, Elsberger B, Going JJ, Orange C, Mallon E, et al. Comparison of Visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. *Br J Cancer.* 2012;106(2):383–8.
  11. Micsik T, Kiszler G, Szabó D, Krecsák L, Hegedűs C, Tibor K, et al. Computer Aided Semi-Automated Evaluation of HER2 Immunodetection—A Robust Solution for Supporting the Accuracy of Anti HER2 Therapy. *Pathol Oncol Res.* 2015;21(4):1005–11.
  12. Hosken B, Coutinho Endringer D, Prandi Campagnaro B, Uggere de Andrade T, Lenz D. High-throughput image analysis in the diagnosis of papillary thyroid carcinoma. *Diagn Cytopathol* [Internet]. julio de 2016;44(7):574–7. Disponible en: <http://doi.wiley.com/10.1002/dc.23481>
  13. Brüggmann A, Eld M, Lelkaitis G, Nielsen S, Grunkin M, Hansen JD, et al. Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains. *Breast Cancer Res Treat.* 2012;132(1):41–9.
  14. Neltner JH, Abner EL, Schmitt FA, Denison SK, Anderson S, Patel E, et al. Digital pathology and image analysis for robust high-throughput quantitative assessment of Alzheimer disease neuropathologic changes. *J Neuropathol Exp Neurol.* 2012;71(12):1075–85.

15. Angell HK, Gray N, Womack C, Pritchard DI, Wilkinson RW, Cumberbatch M. Digital pattern recognition-based image analysis quantifies immune infiltrates in distinct tissue regions of colorectal cancer and identifies a metastatic phenotype. *Br J Cancer* [Internet]. 2013;109(6):1618–24. Disponible en: <http://dx.doi.org/10.1038/bjc.2013.487>
16. Nielsen LAG, Bangsø JA, Lindahl KH, Dahlrot RH, Hjelmberg JVB, Hansen S, et al. Evaluation of the proliferation marker Ki-67 in gliomas: Interobserver variability and digital quantification. *Diagn Pathol*. 2018;13(1):1–8.
17. Helin HO, Tuominen VJ, Ylinen O, Helin HJ, Isola J. Free digital image analysis software helps to resolve equivocal scores in HER2 immunohistochemistry. *Virchows Arch*. 2016;468(2):191–8.
18. Tuominen VJ, Tolonen TT, Isola J. ImmunoMembrane: A publicly available web application for digital image analysis of HER2 immunohistochemistry. *Histopathology*. 2012;60(5):758–67.
19. Papathomas TG, Pucci E, Giordano TJ, Lu H, Duregon E, Volante M, et al. An international Ki67 reproducibility study in adrenal cortical Carcinoma. *Am J Surg Pathol*. 2016;40(4):569–76.
20. Groen R, Abe K, Yoon HS, Li Z, Shen R, Yoshikawa A, et al. Application of microscope-based scanning software (Panoptiq) for the interpretation of cervicovaginal cytology specimens. *Cancer Cytopathol*. 2017;125(12):918–25.
21. Shinde V, Burke KE, Chakravarty A, Fleming M, McDonald AA, Berger A, et al. Applications of Pathology-Assisted Image Analysis of Immunohistochemistry-Based Biomarkers in Oncology. *Vet Pathol*. 2014;51(1):292–303.
22. Achi H El, Belousova T, Chen L, Wahed A, Wang I, Hu Z, et al. Automated diagnosis of

- lymphoma with digital pathology images using deep learning. *Ann Clin Lab Sci.* 2019;49(2):153–60.
23. Sheehan SM, Korstanje R. Automatic glomerular identification and quantification of histological phenotypes using image analysis and machine learning. *Am J Physiol - Ren Physiol.* 2018;315(6):F1644–51.
24. Maity M, Mungle T, Dhane D, Maiti AK, Chakraborty C. An Ensemble Rule Learning Approach for Automated Morphological Classification of Erythrocytes. *J Med Syst.* 2017;41(4).
25. Holtén-Rossing H, Talman M-LM, Jylling AMB, Laenkholm A-V, Kristensson M, Vainer B. Application of automated image analysis reduces the workload of manual screening of sentinel lymph node biopsies in breast cancer. *Histopathology* [Internet]. diciembre de 2017;71(6):866–73. Disponible en: <http://doi.wiley.com/10.1111/his.13305>
26. Qaiser T, Mukherjee A, Reddy PB C, Munugoti SD, Tallam V, Pitkäaho T, et al. HER2 challenge contest: a detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology* [Internet]. enero de 2018;72(2):227–38. Disponible en: <http://doi.wiley.com/10.1111/his.13333>
27. Kent MN, Olsen TG, Feeser TA, Tesno KC, Moad JC, Conroy MP, et al. Diagnostic accuracy of virtual pathology vs traditional microscopy in a large dermatopathology study. *JAMA Dermatology.* 2017;153(12):1285–91.
28. Arvaniti E, Fricker KS, Moret M, Rupp N, Hermanns T, Fankhauser C, et al. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Sci Rep.* 2018;8(1):1–11.
29. Barisoni L, Nast CC, Jennette JC, Hodgins JB, Herzenberg AM, Lemley K V., et al. Digital

- pathology evaluation in the multicenter nephrotic syndrome study network (NEPTUNE). *Clin J Am Soc Nephrol*. 2013;8(8):1449–59.
30. Barisoni L, Hodgin JB. Digital pathology in nephrology clinical trials, research, and pathology practice. *Curr Opin Nephrol Hypertens*. 2017;26(6):450–9.
  31. Waters KM, Cottrell TR, Besharati S, Zhu Q, Anders RA. Evaluation of Peritumoral Fibrosis in Metastatic Colorectal Adenocarcinoma to the Liver Using Digital Image Analysis. *Am J Clin Pathol*. 2019;151(2):226–30.
  32. Gavrielides MA, Conway C, O’Flaherty N, Gallas BD, Hewitt SM. Observer performance in the use of digital and optical microscopy for the interpretation of tissue-based biomarkers. *Anal Cell Pathol*. 2014;2014(2).
  33. Williams BJ, Hanby A, Millican-Slater R, Nijhawan A, Verghese E, Treanor D. Digital pathology for the primary diagnosis of breast histopathological specimens: an innovative validation and concordance study on digital pathology validation and training. *Histopathology*. 2018;72(4):662–71.
  34. Loughrey MB, Bankhead P, Coleman HG, Hagan RS, Craig S, McCorry AMB, et al. Validation of the systematic scoring of immunohistochemically stained tumour tissue microarrays using QuPath digital image analysis. *Histopathology*. 2018;73(2):327–38.
  35. Fuhrich DG, Lessey BA, Savaris RF. Comparison of HSCORE assessment of endometrial beta3 integrin subunit expression with digital HSCORE using computerized image analysis (ImageJ). *Anal Quant Cytopathol Histopathol [Internet]*. agosto de 2013;35(4):210–6. Disponible en: <http://www.ncbi.nlm.nih.gov/pubmed/4090774>
  36. Rizzardi AE, Johnson AT, Vogel RI, Pambuccian SE, Henriksen J, Skubitz APN, et al. Quantitative comparison of immunohistochemical staining measured by digital image

- analysis versus pathologist visual scoring. *Diagn Pathol*. 2012;7(1):1–10.
37. Rizzardi AE, Zhang X, Vogel RI, Kolb S, Geybels MS, Leung YK, et al. Quantitative comparison and reproducibility of pathologist scoring and digital image analysis of estrogen receptor  $\beta$ 2 immunohistochemistry in prostate cancer. *Diagn Pathol* [Internet]. 2016;11(1):1–11. Disponible en: <http://dx.doi.org/10.1186/s13000-016-0511-5>
38. Tollemar V, Tudzarovski N, Boberg E, Törnqvist Andrén A, Al-Adili A, Le Blanc K, et al. Quantitative chromogenic immunohistochemical image analysis in cellprofiler software. *Cytom Part A*. 2018;93(10):1051–9.
39. Kwon AY, Park HY, Hyeon J, Nam SJ, Kim SW, Lee JE, et al. Practical approaches to automated digital image analysis of Ki-67 labeling index in 997 breast carcinomas and causes of discordance with visual assessment. *PLoS One*. 2019;14(2):1–13.
40. Holten-Rossing H, Møller Talman ML, Kristensson M, Vainer B. Optimizing HER2 assessment in breast cancer: application of automated image analysis. *Breast Cancer Res Treat*. 2015;152(2):367–75.
41. Tuominen VJ, Ruotoistenmäki S, Viitanen A, Jumppanen M, Isola J. ImmunoRatio: A publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res*. 2010;12(4).
42. Bean GR, Wen KW, Horvai AE. Adipocyte size variability in benign and malignant lipomatous tumors and morphologic mimics: a quantitative definition using digital pathology. *Hum Pathol* [Internet]. 2018;72:52–8. Disponible en: <https://doi.org/10.1016/j.humpath.2017.10.030>
43. MinCiencias: Biomedical Imaging, Vision and Learning Laboratory. [Internet]. *Ciencia, Tecnología e Innovación en Salud*. Disponible en:

- [https://scienti.minciencias.gov.co/gruplac/jsp/Medicion/graficas/verPerfiles.jsp?id\\_convocatoria=19&nroIdGrupo=00000000003039](https://scienti.minciencias.gov.co/gruplac/jsp/Medicion/graficas/verPerfiles.jsp?id_convocatoria=19&nroIdGrupo=00000000003039)
44. Arvaniti E, Fricker K, Moret M, Rupp N, Hermanns T, Fankhauser C, et al. Replication Data for: Automated Gleason grading of prostate cancer tissue microarrays via deep learning. [Internet]. V1 ed. Harvard Dataverse; 2018. Disponible en: <https://doi.org/10.7910/DVN/OCYCMP>
  45. Kierszenbaum AL, Tres LL. Histología y biología celular. En: Histología y biología celular. 4th editio. Barcelona: Elsevier Health Sciences Spain - T.; 2016. p. 624–7.
  46. Kumar V, Abbas AK, Aster JC, Robbins SL. Robbins basic pathology. Philadelphia, PA.: Elsevier/Saunders; 2013.
  47. Heidenreich A, Bolla M, Joniau S, Mason MD, Matveev V, Mottet N, et al. Guía clínica sobre el cáncer de próstata. 2010;
  48. Hamilton PW, Bankhead P, Wang Y, Hutchinson R, Kieran D, McArt DG, et al. Digital pathology and image analysis in tissue biomarker research. *Methods* [Internet]. 2014;70(1):59–73. Disponible en: <http://dx.doi.org/10.1016/j.ymeth.2014.06.015>
  49. Hipp J, Monaco J, Kunju LP, Cheng J, Yagi Y, Rodriguez-Canales J, et al. Integration of architectural and cytologic driven image algorithms for prostate adenocarcinoma identification. *Anal Cell Pathol*. 2012;35(4):251–65.
  50. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):1–7.
  51. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977;33(1):159.

52. Dere Y, Çelik ÖI, Çelik SY, Ekmekçi S, Evcim G, Pehlivan F, et al. A grading dilemma; Gleason scoring system: Are we sufficiently compatible? A multi center study. *Indian J Pathol Microbiol.* 2020;63(February):S25–9.
53. Ozkan TA, Eruyar AT, Cebeci OO, Memik O, Ozcan L, Kuskonmaz I. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol.* 2016;50(6):420–4.
54. Veloso SG, Lima MF, Salles PG, Berenstein CK, Scalon JD, Bambilra EA. Interobserver agreement of Gleason score and modified Gleason score in needle biopsy and in surgical specimen of prostate cancer. *Int Braz J Urol.* 2007;33(5):639–46.
55. Allsbrook WC, Mangold KA, Johnson MH, Lane RB, Lane CG, Epstein JI. Interobserver reproducibility of Gleason grading of prostatic carcinoma: General pathologists. *Hum Pathol.* 2001;32(1):81–8.

## Apéndices

*Apéndice A. Consentimiento informado aplicado a todos los participantes del presente estudio.*

### **DESEMPEÑO DIAGNÓSTICO Y REPRODUCIBILIDAD INTEROBSERVADOR DEL GRADO DE GLEASON EN LA INTERPRETACIÓN HISTOPATOLÓGICA CONVENCIONAL Y MEDIANTE EL ANÁLISIS DE IMÁGENES POR SOFTWARE**

#### **CONSENTIMIENTO INFORMADO**

Estimado participante:

El Grupo de Investigación PATUIS, del Departamento de Patología de la Universidad Industrial de Santander, se permite solicitar su participación en la investigación titulada “*Desempeño diagnóstico y reproducibilidad interobservador del grado de Gleason en la interpretación histopatológica convencional y mediante el análisis de imágenes por software*”, la cual será realizada por el Dr. Carlos Andrés Sánchez Rueda, médico estudiante de la Especialización en Patología de la misma universidad y miembro del grupo de investigación.

El objetivo de este estudio es establecer la potencial utilidad de un software de análisis de imágenes en la evaluación histopatológica del adenocarcinoma de próstata. A través de los datos obtenidos de los participantes se pretende lograr entrenar a una herramienta informática que será posteriormente usada como apoyo y en conjunción con la evaluación convencional en imágenes de patología digital de adenocarcinoma de próstata.

## DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

Los participantes en el estudio deberán hacer uso de la herramienta QuPath en donde realizarán una demarcación o delimitación de regiones en imágenes histopatológicas que deberán etiquetarse y graduarse según la escala de Gleason. El dataset de imágenes histopatológicas de patología digital se encontrará disponible en un disco virtual, así como apoyos en forma de videos donde se instruye al participante sobre el procedimiento que deberá seguir para lograr la demarcación de las imágenes que serán evaluadas; los vínculos para la descarga de la herramienta QuPath y otros recursos también serán incluidos en el disco virtual mencionado.

La totalidad de las imágenes histopatológicas que serán evaluadas se encuentran divididas en 5 grandes grupos. Los datos obtenidos de los primeros 4 grupos de imágenes serán usados para lograr el entrenamiento de un software de análisis de imágenes que será programado por el grupo de investigación de bioingeniería *Biomedical Imaging, Vision and Learning Laboratory* de la Universidad Industrial de Santander (UIS). Estos primeros 4 grupos de imágenes harán parte de la Fase 1 del estudio. Para el proceso de demarcación de cada uno de estos grupos se permitirá una franja de tiempo variable según el volumen de imágenes que contenga, como se describe en la Tabla 1.

Número del set de imágenes	Número de imágenes del set	Tiempo para la demarcación
Grupo de imágenes 1 (Set 1)	227 imágenes	3 semanas

## DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

Grupo de imágenes 2 (Set 2)	176 imágenes	2 semanas
Grupo de imágenes 3 (Set 3)	105 imágenes	2 semanas
Grupo de imágenes 4 (Set 4)	133 imágenes	2 semanas

Características de los sets de imágenes histopatológicas que se usarán en la primera fase del estudio y el tiempo que tardará su demarcación.

Una vez se cuente con las demarcaciones de los 4 sets mencionados anteriormente, el software de análisis de imágenes desarrollado por el grupo *Biomedical Imaging, Vision and Learning Laboratory* será entrenado en función de estos datos.

En la segunda fase del estudio y en un nuevo grupo de imágenes histopatológicas compuesto por 245 imágenes en total, el profesional etiquetará las imágenes y en cada una de ellas se le indicará una demarcación sugerida por el software de análisis de imágenes. El participante tendrá la posibilidad de modificar su demarcación y estadificación inicial, según prefiera. Ambos patrones definidos por el participante serán sujetos a análisis para evaluar una mayor concordancia con el/los uropatólogo(s) colaborador(es) en el presente estudio (Definido como *Gold Standard* o *Ground Truth* para este trabajo).

## DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

Dentro de las posibles molestias que se considera se podrían llegar a presentar por su participación en este estudio se encuentran cefalea, cansancio y/o mareo, las cuales podrán ser aliviadas con un tiempo de descanso. Los participantes podrán distribuir el tiempo dedicado a las actividades relacionadas con el estudio de la manera que crean más conveniente teniendo en cuenta las fechas límites que se determinen para cada grupo de imágenes.

El investigador y el estudiante de maestría del Grupo de investigación *Biomedical Imaging, Vision and Learning Laboratory*) se encuentran disponibles para resolver cualquier inquietud durante su participación en el presente estudio; al final del presente documento encontrarán los datos de contacto en caso de requerirlos. Usted tiene el derecho a solicitar información acerca de la investigación si así lo desea. Su participación es completamente voluntaria y puede retirarse del estudio cuando lo desee, sin que esto conlleve a ninguna repercusión negativa. No debe cancelar ningún valor monetario por participar en el estudio, así como tampoco se otorgará compensación económica. Su participación satisfactoria en el estudio le significará una mención en los agradecimientos del trabajo final y será tenido en cuenta para posteriores proyectos de la línea de investigación relacionada con el grupo de ingeniería biomédica de la universidad y el departamento de patología.

La información generada por este estudio es estrictamente confidencial y se mantendrá su privacidad, no será utilizada para generar beneficios económicos. Para proteger la identidad de los participantes, al inicio del estudio se le asignará un código mediante el cual será identificado durante todo el proceso. El único que tiene acceso a la información obtenida del estudio será el Ing. Fabián León, quien recibirá los datos de primera mano y establecerá el código que se usará en

## DESEMPEÑO DIAGNÓSTICO EN EL GRADO DE GLEASON

adelante. El investigador no tendrá acceso a la identidad de los participantes. La única variable que se tendrá en cuenta en el proceso será su nivel académico o estudios cursados afines a la oncología, patología quirúrgica o uropatología.

El presente documento se firmará en duplicado quedando una copia en su poder. Los miembros del grupo investigador están en disposición de brindarle cualquier información o pregunta que le surja acerca del estudio. Puede contactarse con la directora del proyecto ó con el investigador en la dirección: Carrera 32 No. 29 – 31 Departamento de Patología, Universidad Industrial de Santander, al correo electrónico o al número telefónico. Para preguntas de carácter técnico respecto al uso de las herramientas informáticas consideradas en este estudio, puede comunicarse con el estudiante de maestría de ingeniería biomédica al correo o al número telefónico. Para preguntas, aclaraciones o inquietudes acerca de los aspectos éticos de esta investigación puede comunicarse con el Comité de Ética para la Investigación Científica de la UIS, en horas hábiles de lunes a viernes de 8-12 y de 2-5pm al teléfono: 6344000 ext. 3208, o enviar correo electrónico a: [comitedetica@uis.edu.co](mailto:comitedetica@uis.edu.co)

Al firmar este consentimiento, acepta que entiende la información que se le ha presentado con respecto a beneficios, riesgos y que está de acuerdo en autorizar la participación en este estudio.

Con fecha \_\_\_\_\_, habiendo comprendido lo anterior y una vez que se le aclararon todas las dudas que surgieron con respecto a su participación en la investigación, usted acepta participar en investigación titulada:

**DESEMPEÑO DIAGNÓSTICO Y REPRODUCIBILIDAD INTEROBSERVADOR DEL GRADO DE GLEASON EN LA INTERPRETACIÓN HISTOPATOLÓGICA CONVENCIONAL Y MEDIANTE EL ANÁLISIS DE IMÁGENES POR SOFTWARE**

---

Nombre del participante

Firma

Cédula

¿Cursó usted un *fellowship* en uropatología? \_\_\_\_ (Sí/No)

¿Cursó usted un *fellowship* en patología quirúrgica? \_\_\_\_ (Sí/No)

¿Cursó usted un *fellowship* en patología oncológica? \_\_\_\_ (Sí/No)

Responsable de toma de consentimiento informado

Carlos Andrés Sánchez Rueda

Médico Residente Patología UIS

*Apéndice B. Modelo de la invitación para la convocatoria a la participación del estudio.*

**El Grupo de Investigación PATUIS  
del Departamento de Patología**

Tiene el gusto de convocar a médicos residentes de patología, médicos patólogos y uropatólogos a participar en la investigación titulada:

**“Desempeño Diagnóstico y Reproducibilidad Interobservador del Grado de Gleason en la Interpretación Histopatológica Convencional y Mediante el Análisis de Imágenes por Software”**

**GRUPO DE INVESTIGACIÓN  
PATUIS**  
Patología Estructural, Funcional y  
Clínica de la UIS

**ALIANZA CON**  
**Grupo de Investigación BV2Lab**  
Biomedical Imaging, Vision and  
Learning Laboratory

Autor:  
**Dr. Carlos Andrés Sánchez Rueda**  
Médico residente de primer año

Fecha:  
Hora:  
Plataforma:  
ID:

**DIRECTORA**  
Olga Mercedes Álvarez Ojeda.  
MD Patóloga

**CODIRECTORA**  
Tania Mendoza Herrera. MD  
Internista. MSc Epidemiología

**Universidad Industrial de Santander**

**72**  
Universidad Industrial de Santander  
Patrimonio educativo y cultural

#LaUISqueQueremos