



Detección de cáncer de próstata en imágenes microscópicas de biopsias

Carlos Fernando Quintero Olaya
Jhonier Andrés Vargas Gómez

Escuela de Ingeniería Eléctrica y Electrónica
Facultad de Ingeniería
Universidad del Valle
Santiago de Cali
2023



Detección de cáncer de próstata en imágenes microscópicas de biopsias

Carlos Fernando Quintero Olaya

Jhonier Andrés Vargas Gómez

Trabajo presentado para optar por el título de ingeniero electrónico

Director

Humberto Loaiza Correa, PhD

Grupo de investigación Percepción y Sistemas Inteligentes

Escuela de Ingeniería Eléctrica y Electrónica

Facultad de Ingeniería

Universidad del Valle

Santiago de Cali

2023

Tabla de contenido

Nota de aceptación	ix
Resumen.....	xi
Abstract.....	xiii
1. Introducción	17
1.1. Contextualización.....	17
1.2. Planteamiento del problema y objetivos	18
1.2.1. Planteamiento del problema.....	18
1.2.2. Objetivo general	18
1.2.3. Objetivos específicos	19
1.3. Descripción del trabajo	19
1.4. Organización de los capítulos	20
2. Antecedentes y marco teórico	20
2.1. Antecedentes	21
2.1.1. Antecedentes nacionales	21
2.1.2. Antecedentes internacionales	21
2.1.3. Discusión.....	26
2.2. Marco teórico	29
2.2.1. Cáncer de próstata.....	29
2.2.1.1. Adenocarcinoma.....	29
2.2.1.2. Biopsia	29
2.2.1.3. Coloración H&E	30
2.2.1.4. Clasificación Gleason	30
2.2.1.5. Micrografía.....	30
2.2.2. Preprocesamiento de imágenes.....	31
2.2.2.1. Preprocesamiento en escala de grises	32
2.2.2.1.1. Corrección Gamma.....	32
2.2.2.1.2. Transformada Wavelet multi nivel	32
2.2.2.2. Preprocesamiento a color	33
2.2.2.2.1. Normalización de la tinción	33
2.2.3. Segmentación de imágenes	33

2.2.3.1.	Segmentación método Otsu	33
2.2.3.2.	Segmentación por contornos activos	33
2.2.3.3.	Segmentación por método Watershed.....	33
2.2.3.4.	Segmentación basada en agrupamiento – clustering	34
2.2.3.5.	Métricas de evaluación	34
2.2.3.5.1.	Índice de Jaccard	34
2.2.3.5.2.	Índice de Dice	34
2.2.3.5.3.	Índice BF score	34
2.2.4.	Extracción de características.....	34
2.2.4.1.	Características por contorno	35
2.2.4.2.	Características por textura y momentos estadísticos.....	35
2.2.4.2.1.	Matriz GLCM.....	35
2.2.4.2.2.	Estadísticas por canal R, G y B	36
2.2.5.	Reconocimiento de patrones.....	36
2.2.5.1.	Características.....	36
2.2.5.2.	Tipos de aprendizaje.....	36
2.2.6.	Máquinas de aprendizaje superficial	37
2.2.6.1.	Naive Bayes.....	37
2.2.6.2.	k-Nearest Neighbors	37
2.2.6.3.	Support Vector Machine	37
2.2.6.4.	Multilayer Perceptron	37
2.2.7.	Máquinas de aprendizaje profundo	38
2.2.7.1.	AlexNet	38
2.2.7.2.	ResNet-50.....	38
2.2.7.3.	EfficientNetB0	39
2.2.8.	Métricas de evaluación.....	40
2.2.8.1.	Precisión	40
2.2.8.2.	Sensibilidad	40
2.2.8.3.	F1 score	40
2.2.8.4.	Exactitud	40
2.2.9.	Discusión.....	40
3.	Descripción del dataset	41

4. Metodología de la solución propuesta: Software	45
4.1. Introducción	45
4.2. Preprocesamiento	47
4.2.1. Obtención de los parches	47
4.2.2. Conversión a escala de grises	50
4.2.2.1. Corrección gamma	50
4.2.2.2. Transformada Wavelet	50
4.2.3. Preprocesamiento a color	51
4.3. Segmentación	52
4.3.1. Segmentación basado en umbrales y regiones	52
4.3.2. Segmentación basada en colores por clustering	52
4.4. Extracción de características	53
4.4.1. Por contorno	53
4.4.2. Por textura y momentos estadísticos	53
4.4.3. Selección de características	54
4.5. Clasificación	55
4.5.1. Clasificadores - Máquinas superficiales	55
4.5.2. Clasificadores - Máquinas profundas	56
4.6. Evaluación del tejido completo	57
4.7. Conclusiones	57
5. Pruebas y resultados	59
5.1. Introducción	59
5.2. Máquinas de aprendizaje superficial	59
5.2.1. Descripción	59
5.2.2. Procedimiento: Imágenes con tamaño de 256x256 pixeles	59
5.2.3. Resultados y análisis	60
5.2.4. Procedimiento: Imágenes de tamaño 512x512 pixeles	61
5.2.5. Resultados y análisis	62
5.3. Pruebas – Máquinas profundas	63
5.3.1. Descripción	63
5.3.2. Procedimiento: Imágenes con tamaño de 256x256 pixeles	63
5.3.3. Resultados y análisis	63

5.3.4.	Procedimiento: Imágenes con tamaño de 512x512 pixeles	65
5.3.5.	Resultados y análisis.....	65
5.4.	Diagnóstico por tejidos – Análisis Comparativo	66
5.4.1.	Descripción.....	66
5.4.2.	Procedimiento	66
5.4.3.	Resultados y análisis.....	66
5.5.	Alcances	68
5.6.	Limitaciones	69
5.7.	Conclusiones	70
6.	Conclusiones generales	72
7.	Trabajos futuros.....	75
8.	Referencias bibliográficas	77
9.	Anexos	82

Lista de figuras

Ilustración 2-1. Puntuación Gleason [23]	30
Ilustración 2-2. Partes que se pueden encontrar en una imagen de biopsia de próstata.....	31
Ilustración 2-3. Arquitectura red AlexNet tomada de [37].....	38
Ilustración 2-4. Ejemplo de conexión residual, tomado de [38]	39
Ilustración 2-5. a) Modelo de CNN convencional, b) – d) Modelos que solo tratan una dimensión en las imágenes, e) Modelo propuesto en [39].	39
Ilustración 3-1. Tejidos con su respectiva máscara según la gravedad completa del tejido	43
Ilustración 3-2. Gravedad del Tejido de Grado 2, con su tejido y máscara	44
Ilustración 3-3. Parches con diferentes puntuaciones Gleason	44
Ilustración 4-1. Diagrama de bloques metodología de solución con máquinas superficiales-clasificación por parches.....	46
Ilustración 4-2. Diagrama de bloques metodológica de solución – clasificación tejido completo.	46
Ilustración 4-3. Imagen WSI de ejemplo de la base de datos que se dispone.	47
Ilustración 4-4. Image patching de las imágenes del dataset para una mejor representación de la información.....	48
Ilustración 4-5. Presencia de tinción no deseada en los tejidos.....	49
Ilustración 4-6. Efecto de aplicar conversión escala de grises y corrección Gamma.....	50
Ilustración 4-7. Aplicación de la transformada Wavelet multi nivel a las imágenes.....	51
Ilustración 4-8. Manipulación del color en los canales H y E.	51
Ilustración 4-9. Aplicación del algoritmo k means en las imágenes.	53
Ilustración 4-10. Pruebas realizadas con las máquinas superficiales.	55
Ilustración 4-11. Diagrama de bloques de la prueba 5.....	56
Ilustración 4-12. Clasificación a través de las máquinas profundas.....	57
Ilustración 5-1. Solución tomada para la evaluación y extracción de características de las imágenes 512 x 512.....	61

Lista de Tablas

Tabla 2-1. Resultado de clasificación transformada Wavelet nivel 1 [13]	23
Tabla 2-2. Resultado de clasificación transformada Wavelet nivel 2 [13]	24
Tabla 2-3. Resultado de desempeño en [14]	24
Tabla 2-4. Resumen de los antecedentes.....	28
Tabla 3-1. Contenido del dataset escogido.....	41
Tabla 3-2. Contenido set de imágenes de la universidad de Radboud.....	42
Tabla 3-3. Contenido set de imágenes del instituto de Karolinska.....	44
Tabla 4-1. Intervalo de valores para selección de parches sin ruido por tinción no deseada.	49
Tabla 4-2. Índices de calidad de la segmentación de los 3 métodos de segmentación.....	52
Tabla 4-3. Características de tipo morfológicas extraídas.	53
Tabla 4-4. Características de tipo textura extraídas a las imágenes.....	54
Tabla 4-5. Características estadísticas extraídas por canal R, G y B.	54
Tabla 5-1. Resultados máquinas superficiales imágenes de 256x256 pixeles	60
Tabla 5-2. Prueba 2 máquinas superficiales tamaño de imágenes 512x512.	62
Tabla 5-3. Métricas desempeños máquinas superficiales.....	62
Tabla 5-4. Hiperparámetros clasificador SVM.....	62
Tabla 5-5. Hiperarámetros clasificador MLP.....	63
Tabla 5-6. Resultados clasificación multiclase máquinas profundas con imágenes 256x256.	64
Tabla 5-7. Métricas de desempeño por clase.	64
Tabla 5-8. Resultados máquinas profundas imágenes de 512x512 pixeles.	65
Tabla 5-9. Métricas de desempeños redes profundas.	65
Tabla 5-10. Resultados evaluación del tejido completo.	67
Tabla 5-11. Métricas por clase clasificación tejido completo.....	67
Tabla 5-12. Hiperpárametros máquinas MLP diagnóstico tejido completo.	68

Nota de aceptación

Director

Jurado

Jurado

Agradecimientos

Durante el desarrollo de nuestro trabajo de grado “Detección de cáncer de próstata en imágenes microscópicas de biopsias” nos ha representado un logro, en el que hemos pasado un camino largo y arduo, lleno de muchos conocimientos del que hemos aprendido y ejecutado para completarlo. Siendo así, tenemos las siguientes palabras de agradecimientos y reflexiones personales que nos han dejado a cada uno:

Este trabajo de grado ha representado, para mí, Carlos Fernando Quintero, un trayecto duro y agotador debido a los diferentes problemas que ocurrían en la consecución de este trabajo, trayendo consigo desánimo y un sentimiento pesimista, que no ayudaba a la terminación del documento. De tal modo, que gracias a la esperanza y vida que me brindó Dios, representó un apoyo emocional que junto a su lado pude resolver los problemas, en el dónde me hizo entender que el trabajo arduo y persistencia es importante principalmente en el área de visión artificial y reconocimiento de patrones.

Por otro lado, quiero agradecer a mis Padres, por su apoyo, amor brindado y ánimos dados, siendo ellos la parte fundamental por la que seguí desarrollando este trabajo. A su vez, también quiero agradecer a las enseñanzas brindadas por mi hermano, David Alfredo Quintero, ya que al ser un Ingeniero Electrónico, él ha sido siempre una pieza fundamental dándome los consejos necesarios que necesitaba durante mi desarrollo académico como profesional.

Por último, y más importante, quiero agradecer a Humberto Loaiza por ser nuestro director, y ayuda fundamental en la consecución del trabajo, brindándonos siempre diferentes espacios que nos aportaban en la solución del proyecto, siendo necesarias sus clases de Visión Artificial y Reconocimiento de Patrones ya que, con estas, nos brindaron el conocimiento necesario para la misma solución del trabajo.

Del mismo modo, para mí, Jhonier Vargas, este trabajo de grado ha representado un conjunto de emociones tanto buenas como regulares e incluso malas en ocasiones, ha representado un reto para mí desde darme cuenta que este campo de la visión artificial, reconocimiento de patrones, análisis de datos y ciencia de datos ha significado para mí que es un sinfín de aprendizaje y que hay que trabajar fuerte para tener un dominio completo y consciente y es por esto que expreso mi agradecimiento por todo el apoyo que recibí a través del proceso.

Quiero agradecer al profesor Humberto Loaiza por aceptarnos y brindarnos un tema para desarrollar como trabajo de grado, a mi compañero Carlos Fernando Quintero Olaya por haber aceptado ser mi compañero de trabajo de grado y estar en todos los momentos tanto los buenos y no tan buenos, a mí madre y abuelo por estar ahí presentes y ser un apoyo emocional, a mis compañeros de la universidad y a los profesores que aportaron desde sus conocimiento a la formación de este estudiante y que intentó siempre traducir esas enseñanzas en un buen proceder e intento de hacer las cosas de la manera correcta.

Resumen

En este documento se encuentra registrado la implementación de diferentes modelos de clasificación entre máquinas superficiales y profundas para la clasificación de la severidad del cáncer de próstata en imágenes histopatológicas de biopsias del tejido prostático por medio de técnicas de visión artificial y reconocimiento de patrones. La severidad de afectación del tejido prostático en una imagen se graduó en una escala de 0-5, con base en los resultados de la clasificación de múltiples parches extraídos (Image Patching). Cada parche se clasificó de acuerdo con el sistema de puntuación de Gleason: puntuación Gleason 0 para una zona del tejido que es benigno, Gleason 3, 4 y 5, siendo 3 el menos severo y 5 el más severo.

Este trabajo no pretende reemplazar al especialista, sino que tiene como propósito brindar herramientas de apoyo por medio de sistemas de diagnóstico asistido por computadora (CAD). Estas herramientas buscan generar alertas e información sobre condiciones del tejido en la imagen de manera muy rápida para que el especialista pueda disponer de más tiempo para realizar diagnósticos complejos, establecer procedimientos y realizar seguimientos a los pacientes.

La implementación de las máquinas superficiales involucra la aplicación de técnicas de visión artificial como conversión a escala de grises, corrección gamma, descomposición y conversión del modelo de color RGB por canales y al modelo HSV respectivamente. El tipo de características que se extrajeron de las imágenes fue de tipo textura, a través de la matriz GLCM y características de tipo estadísticas a través de hacer la descomposición del modelo RGB. Para las máquinas profundas se implementó transfer learning, una técnica con el fin de evitar sobre entrenamiento (overfitting), pues se usaron modelos previamente entrenados en datasets de imágenes para clasificación multiclase.

En el desarrollo del trabajo se usaron las máquinas superficiales que mayor desempeño presentaron las cuales fueron SVM y MLP y a su vez se usaron las máquinas profundas ResNet50 y EfficientNetB0 para una clasificación multiclase de parches individuales y a su vez fueron puestos a prueba en el diagnóstico de la severidad del cáncer de próstata de un tejido completo (escala de 0-5).

Para SVM presentó un desempeño en precisión, sensibilidad, F1 score y exactitud en la clasificación de parches individuales de 0.750, 0.753, 0.744 y 0.753, pero para el diagnóstico del tejido completo fue de 0.533, 0.542, 0.536 y 0.542.

Para MLP presentó un desempeño en precisión, sensibilidad, F1 score y exactitud en la clasificación de parches individuales de 0.774, 0.773, 0.773 y 0.773, pero para el diagnóstico del tejido completo fue de 0.513, 0.536, 0.521 y 0.536.

Para ResNet50 presentó un desempeño en precisión, sensibilidad, F1 score y exactitud en la clasificación de parches individuales de 0.862, 0.859, 0.858, 0.859, para el diagnóstico del tejido completo fue de 0.954 en todas las métricas.

Para EfficientNetB0 presentó un desempeño en precisión, sensibilidad, F1 score y exactitud en la clasificación de parches individuales de 0.844, 0.844, 0.843 y 0.844, para el diagnóstico del tejido completo fue de 0.958 y 0.957 para el resto de las métricas.

Palabras clave: Máquinas superficiales, máquinas profundas, grado Gleason, sistemas CAD, visión artificial, reconocimiento de patrones

Abstract

This paper reports the implementation of different classification models between superficial and deep machines for the classification of prostate cancer severity in histopathological images of prostate tissue biopsies by means of computer vision and pattern recognition techniques. The severity of prostate tissue involvement in an image was graded on a scale of 0-5, based on the results of the classification of multiple extracted patches (Image Patching). Each patch was classified according to the Gleason scoring system: Gleason score 0 for an area of tissue that is benign, Gleason 3, 4 and 5, with 3 being the least severe and 5 the most severe.

This work is not intended to replace the specialist, but rather to provide support tools by means of computer-aided diagnosis (CAD) systems. These tools seek to generate alerts and information on tissue conditions in the image very quickly so that the specialist can have more time to make complex diagnoses, establish procedures and follow up patients.

The implementation of the surface machines involves the application of computer vision techniques such as grayscale conversion, gamma correction, decomposition, and conversion of the RGB color model by channels and to the HSV model respectively. The type of features extracted from the images were texture type, through the GLCM matrix and statistical type features through doing the decomposition of the RGB model. For the deep machines, transfer learning was implemented, a technique to avoid overfitting, since models previously trained on image datasets were used for multiclass classification.

In the development of the work, the surface machines with the best performance were used, which were SVM and MLP, and in turn the deep machines ResNet50 and EfficientNetB0 were used for a multiclass classification of individual patches and in turn were tested in the diagnosis of prostate cancer severity of a complete tissue (scale of 0-5).

For SVM it presented a performance in precision, sensitivity, F1 score and accuracy in the classification of individual patches of 0.750, 0.753, 0.744 and 0.753, but for the diagnosis of whole tissue was 0.533, 0.542, 0.536 and 0.542.

For MLP it presented a performance in precision, sensitivity, F1 score and accuracy in the classification of individual patches of 0.774, 0.773, 0.773 and 0.773, but for the diagnosis of the complete tissue was 0.513, 0.536, 0.521 and 0.536.

For ResNet50 it presented a performance in precision, sensitivity, F1 score and accuracy in individual patch classification of 0.862, 0.859, 0.858, 0.859, for whole tissue diagnosis was 0.954 in all metrics.

For EfficientNetB0 it presented a performance in precision, sensitivity, F1 score and accuracy in the classification of individual patches of 0.844, 0.844, 0.843 and 0.844, for the diagnosis of the whole tissue was 0.958 and 0.957 for the rest of the metrics.

Key words: Shallow machines, deep machines, Gleason grade, CAD systems, machine vision, pattern recognition

1. Introducción

La mayoría de los casos de cáncer de próstata pertenecen a la categoría de adenocarcinoma, siendo una enfermedad que afecta a las células que forman glándulas. Este tipo de cáncer afecta a la próstata, una glándula del sistema reproductor masculino, esto se da debido al crecimiento sin control de las células glandulares en la próstata [1]. Es la cuarta enfermedad más diagnosticada en el mundo, para el 2020 se encontraron que 1.414.259 personas en el mundo fueron diagnosticadas [2]. En Colombia es una de las enfermedades prevalentes reportadas, con una tasa de mortalidad de 13.76 por 100000 habitantes en el 2020 [3].

El diagnóstico final de la presencia o no del cáncer de próstata en el paciente se da por medio de una biopsia, procedimiento que consiste en retirar tejido de la próstata que es examinado luego por el especialista autorizado para dar el veredicto respecto la presencia del cáncer, así como la severidad del cáncer por medio de la puntuación de Gleason [4]. Proceso que es subjetivo pues es realizado por el humano y que es susceptible a errores humanos; en donde se ha evidenciado diferencias entre la asignación del grado de severidad de un mismo diagnóstico por diferentes patólogos o especialistas (variabilidad interobservador e intraobservador subóptima) con un desacuerdo entre diagnósticos del 30% al 53% [5], [6].

En base a esto se buscó implementar diferentes máquinas de aprendizaje superficial y profundo con el fin brindar una solución por software para la clasificación automática del diagnóstico de este tipo de imágenes, que sea lo más confiable posible y no genere duda a confusiones entre los diferentes grados.

De esta manera se registra este documento las diferentes discusiones con relación a lo encontrado en los antecedentes o trabajos previos realizados en base a ciertas técnicas, con distintos enfoques y propuestas de solución que sirvieron de punto de partida para así tener una propuesta de solución con una metodología para abordar el diagnóstico del cáncer de próstata. Se registra también apartados de alcances, limitaciones y conclusiones de los resultados obtenidos.

1.1. Contextualización

Este es un trabajo pensado y realizado con el fin de aportar desde el campo de la visión artificial y reconocimiento de patrones un sistema de diagnóstico asistido por computador (CAD) [7], al campo de la medicina, específicamente al diagnóstico de la severidad del cáncer de próstata. Es importante hacer mención de que este es un sistema que no pretende reemplazar al especialista pues es quién tiene autoridad y reconocimiento oficial para dar un diagnóstico, pero teniendo en cuenta que el humano es susceptible a errores, es por lo que se busca implementar un sistema que ayude a un diagnóstico preliminar. Para esta implementación se tiene en cuenta la clasificación Gleason, que es una puntuación dada por los especialistas, basada en la estructura del tejido celular de la próstata en función de su capacidad de mantener un patrón estructural ordenado, donde no sea evidente una división celular excesiva que indique presencia de cáncer y además el patrón en la forma de las glándulas es clave para diferenciar de un grado a otro, donde además es posible tener patrones de cáncer mayoritario y minoritario la cual asigna una puntuación de Gleason 0, 3,

4 y 5. Donde Gleason 0 se considera una zona donde el tejido es benigno y de 3 a 5 diferente severidad del cáncer de próstata siendo 3 el menor y 5 el de mayor severidad. En donde, al evaluar diferentes zonas del tejido permitirá así obtener una valoración de la gravedad del tejido en una escala de 0-5.

1.2. Planteamiento del problema y objetivos

1.2.1. Planteamiento del problema

Las biopsias de la próstata se realizan a pacientes con sospechas de tener cáncer de próstata una vez se ha realizado los diagnósticos DRE y/o PSA [4]. La biopsia es un procedimiento del cual se extraen cilindros de tejido prostático por medio de una aguja que ingresa a la próstata del paciente. Estos cilindros de tejido se cortan, se toman imágenes de estos cortes, en donde se determina si hay o no tejido anormal que indique la presencia de cáncer; la gravedad de haber células cancerígenas en este tejido se clasifica en grupos de 0 a 5 en función de la puntuación de Gleason [4].

La determinación del grupo o gravedad del cáncer del tejido es llevada a cabo por un especialista que tiene por trabajo analizar cada imagen, en donde las muestras presentan más de un tipo de tejido que debe tener en cuenta para la clasificación del grupo, además las imágenes para analizar pueden provenir no solo de 1 muestra, sino de 10 o más muestras que luego se convierten en un número considerable de muestras para analizar por paciente.

Es por esto por lo que se busca desde el área del procesamiento digital de señales aportar al análisis y determinación de las imágenes microscópicas para la detección del cáncer de próstata.

Teniendo en cuenta lo anteriormente mencionado, se genera la siguiente pregunta: ¿Cómo detectar el cáncer de próstata y su gravedad en imágenes digitales de biopsias utilizando técnicas de visión artificial y máquinas de aprendizaje superficial y profundas? Obteniendo como resultado diferentes clasificadores con máquinas de aprendizaje superficiales y profundo, los cuales generaron diferentes desempeños a través de la aplicación de diferentes técnicas de visión artificial y reconocimiento de patrones, así como también la implementación de máquinas de aprendizaje profundo. El trabajo se llevó a cabo partiendo del conjunto de imágenes proporcionado por el centro médico Radboud, siendo accesible a través del sitio web de Kaggle [8].

Con este trabajo no se pretende reemplazar al especialista, quién es la persona autorizada y certificada para dar el diagnóstico final al paciente. Lo que se busca es aportar desde la ingeniería electrónica y el campo de procesamiento de imágenes, visión artificial y reconocimiento de patrones al campo de la salud una ayuda al momento de la detección del cáncer de próstata.

1.2.2. Objetivo general

Detectar presencia y el grado de Gleason del adenocarcinoma en imágenes microscópicas de biopsias de próstatas mediante el uso de técnicas de visión artificial y reconocimiento de patrones.

1.2.3. Objetivos específicos

1. Determinar el ajuste y escalamiento adecuado para las imágenes de biopsias mediante técnicas de procesamiento de imágenes, resaltando así las regiones con presencia o ausencia del adenocarcinoma de próstata.
2. Extraer las características presentes en imágenes microscopias de biopsias que sirvan en la detección y determinación del grado Gleason del adenocarcinoma de próstata.
3. Implementar dos máquinas de aprendizaje superficial para la clasificación de la presencia y grado Gleason del adenocarcinoma de próstata a partir de las características extraídas de las imágenes microscópicas de biopsias.
4. Implementar dos máquinas de aprendizaje profundo para la clasificación de la presencia y grado Gleason del adenocarcinoma de próstata a partir de las imágenes microscópicas de biopsias.
5. Evaluar el desempeño de las máquinas de aprendizaje implementadas.

1.3. Descripción del trabajo

El desarrollo del sistema de detección y determinación de la gravedad de cáncer en imágenes digitales de biopsias de próstata utiliza técnicas de visión artificial y máquinas de aprendizaje superficial (Shallow Learning) y profundo (Deep Learning). El sistema se concibió con base en la configuración de los sistemas de Diagnóstico Asistido por Computador (Computer Assisted Diagnosis, CAD).

El sistema que incluye máquinas de aprendizajes superficiales tiene una etapa de procesamiento para la adecuación de las imágenes (Conversión de color, modelo de color), reducción de ruido como coloraciones no deseadas por tinción química, eliminación de datos atípicos, aumento del contraste (Corrección Gamma, Normalización del color en base a la tinción) y segmentación (Método Otsu, Watershed y Contornos activos), que tienen como finalidad potenciar la detección y diferenciación de las zonas de interés. Sobre las regiones segmentadas, se extraen los vectores de características (GLCM, morfológicas, estadísticas), con mayor poder discriminante entre las clases. Estas características constituyen la información para entrenar y probar las máquinas de aprendizaje Naive Bayes, kNN, SVM y MLP.

El sistema con las máquinas de aprendizaje profundo (Deep Learning) incluye internamente el proceso de extracción de características y clasificación, y recibe como entrada las imágenes (en lugar de vectores de características). Por contener muchas capas, requiere de una mayor base de datos y algoritmos más complejos para la sintonización de sus parámetros. Las redes usadas fueron AlexNet, ResNet50 y EfficientNetB0, todas estas redes poseen capas para evitar el sobre ajuste (Capas de dropout y técnicas de regularización), a los datos y así tener un nivel de generalización sobre los datos que aprenden. Además, poseen diferentes arquitecturas bajo las cuales fueron concebidas para presentar mejor desempeño en las tareas de clasificación de imágenes.

En base a esto la clasificación se lleva a cabo a través de una clasificación multiclase de parches obtenidos de cada imagen (Image Patching) con diagnóstico por parches y la clasificación de una imagen total dando diagnóstico sobre esta en su totalidad.

1.4. Organización de los capítulos

El documento se divide en 6 capítulos. En el primero se expone la contextualización el problema encontrado, solución que se le quiere dar y cómo queremos abordar dicho problema encontrado. En el capítulo 2 se presenta la bibliografía encontrada respecto a los trabajos hechos de sistemas CAD para la clasificación automática de imágenes histopatológicas, en este caso biopsias de próstata para la detección de la severidad de cáncer de próstata. En el capítulo 3, se explica la solución propuesta, así como la metodología de solución, en donde se aborda más a fondo lo realizado en cada etapa de los sistemas CAD. En el capítulo 4, se registran los resultados obtenidos de los diferentes sistemas constituidos a partir de lo realizado en cada etapa, el procedimiento, resultados y análisis de los resultados. Para el capítulo 5, se expresan las conclusiones obtenidas después de haber hecho el ejercicio de clasificación y finalmente en el capítulo 6 se aborda qué planes a futuros quedan expuestos para llevar a cabo con base en lo que se logró y no logró y puede aportar a un mejoramiento de los resultados.

2. Antecedentes y marco teórico

Se presentan información sobre trabajos similares y afines sobre sistemas de clasificación de patologías en imágenes diagnósticas que empleen máquinas de aprendizaje superficial y profundo. En particular se hace énfasis en las metodologías y técnicas de visión artificial y reconocimiento de patrones utilizadas para la detección y clasificación del cáncer de próstata. Se consultaron las bases de datos IEEE, Scopus, y Science Direct. En donde se tuvo en cuenta los criterios de inclusión y exclusión como búsqueda entre los años 2016 y 2023, utilizando las palabras claves: Cáncer de Próstata, Clasificación de Gleason, Procesamiento de Imágenes, Reconocimiento de Patrones, Aprendizaje Superficial, Aprendizaje Profundo. A su vez se tuvo en cuenta artículos más citados, trabajos con enfoque en clasificación de imágenes histopatológicas, trabajos con el sistema de puntuación Gleason y el área de ciencias de la computación.

De acuerdo con estos criterios se obtuvo un total de 834 documentos, en 2016 se encuentran 41 documentos publicados y 160 para el año 2023, lo que indica un aumento en publicaciones e interés en el tema con el paso del tiempo desde el 2016 hasta el 2023. Cabe resaltar que el país con más investigaciones y documentos publicados es Estados Unidos en primer lugar, China en segundo lugar y Canadá en tercer lugar, también se encuentran países como India, Italia, Reino Unido, Corea del Sur, España, Alemania y Australia. También se encontró que The University of British Columbia es la institución con más publicaciones teniendo 30 documentos, seguida por Queen's University y National Institutes of Health con 22 y 18 documentos respectivamente.

También se incluye el marco teórico con la explicación de los conceptos requeridos para este trabajo de grado, desde la definición del grado Gleason, las técnicas usadas para

procesamiento de imágenes y las máquinas de aprendizaje superficial y profundo que se utilizan en la tarea de clasificación de imágenes histopatológicas.

2.1. Antecedentes

El reconocimiento de enfermedades genéticas como el cáncer han sido objeto de estudio y de investigación tal que existen campos enfocados o que tienen interés a este tipo de clasificación ya sea como los sistemas CAD y el análisis de imágenes histopatológicas, pues da pie a probar técnicas de visión artificial y reconocimiento de patrones en conjunto para reconocer hasta qué punto puede llegar a ser un algoritmo más eficiente que otro, qué características aportan más a la tarea de discriminación entre tipos de severidad de un cáncer cuando de un diagnóstico apropiado se trata, pues dependiendo de esto el paciente tendrá o no chances de un tratamiento efectivo y una mejor calidad de vida respecto al cuidado de su salud.

2.1.1. Antecedentes nacionales

Towards clinical significance prediction using k trans evidences in prostate cancer [9].

En [9] se encuentra registrado el trabajo llevado a cabo en la detección de cáncer de próstata bajo el uso de imágenes del tipo k-trans pertenecientes a un data set público de 339 imágenes de este tipo tomadas de The Cancer Imaging Archive (TCIA). La detección se hizo bajo el uso de 3 algoritmos diferentes, k-Nearest Neighbors (k-NN), Support Vector Machine (SVM) y Random Forest (RaF) en 3 diferentes regiones, en la periférica, transicional y anterior de la región del estroma fibromuscular. Se dispuso 60% del contenido del data set para entrenamiento y 40% para pruebas. Los resultados obtenidos en términos de precisión de presencia del cáncer de próstata de los diferentes algoritmos fueron de 80.83% para SVM, 76.83% para k-NN y 75.89% para RaF. De este documento cabe destacar que su objetivo fue el de hacer una detección binaria del cáncer de próstata, es decir, se centraron en que los algoritmos pudieran detectar si había o no presencia del cáncer de próstata, pero no llevaron a cabo el tratamiento de los algoritmos para la distinción de los diferentes grados del adenocarcinoma de próstata en base de la puntuación Gleason.

2.1.2. Antecedentes internacionales

Towards grading Gleason score using generically trained deep convolutional neural networks [10].

En [10] el trabajo realizado consiste en 3 partes, primero usa una CNN pre entrenada llamada OverFeat, a la cual le dan uso como extractor de características, es decir, usan ciertas capas para extraer características a un nivel más o menos profundo y las capas finales de clasificación las ignoran. Específicamente usan las capas 9 y 16 para extraer las características, además usan dos niveles de aumento, 5x y 10x. Segundo, llevan a cabo la clasificación por parches usando Random Forest y SVM para los dos tipos de aumento y las características extraídas de ambas capas y en tercer lugar hacen la clasificación de la imagen entera, es decir, realizan el recorte de la imagen completa y someten a clasificación el conjunto de parches y función del grado que haya sido mayor clasificado entre los 4 (Grado

Gleason benigno 0, 3, 4 y 5), se hace votación y se realiza el diagnóstico del grado con mayor voto o el que fue más clasificado.

Como resultados para la clasificación multiclase de los parches logran 81.1% y 79.3% para SVM y Random Forest respectivamente y para la clasificación del tejido o imagen completos como lo presentan ellos, obtienen 89.2% y 73.2% para SVM y Random Forest respectivamente.

A Review on Prediction and Prognosis of the Prostate Cancer and Gleason Grading of Prostatic Carcinoma Using Deep Transfer Learning Based Approaches [11]

En el artículo [11], se centra en el desarrollo de máquinas profundas encargadas de detectar la agresividad del cáncer de próstata mediante la extracción automática de características de imágenes de diapositivas completas de biopsias de próstata, las cuales fueron brindadas por los centros Médicos de Karolinska y Radboud. Durante la investigación resalta que utilizaron 10616 imágenes de tejidos completos con diferentes 6 grados de severidad de cáncer, que están en una escala de 0-5. Estas imágenes varían en sus dimensiones, remarcando que es necesario hacer un reescalamiento de todas las imágenes a una dimensión de 224x224 píxeles. Por otra parte, destacan el uso de máquinas como MobileNet V2, InceptionResNet V2, DenseNet 169, ResNet101 V2 y NASNetMobile, en donde obtuvieron una exactitud para las distintas máquinas de 80.18%, 84.99%, 83.98%, 81.50%, 75.13%, respectivamente. Cabe destacar, que la máquina DenseNet 169 es la mejor máquina en saber detectar la severidad del cáncer con Grado 0, Grado 2 y Grado 3, con un porcentaje de exactitud de 89.76%, 96.98% y 91.98%, para estos mismos grados de severidad del cáncer. Por otro lado, las máquinas que mejor clasifican el Grado 1, Grado 4, Grado 5, son las máquinas ResNet 101 V2, NASNetMobile, MobileNet V2, con una exactitud de 79.43%, 99.66%, 99.84%, en este mismo orden.

Computerized Classification of Prostate Cancer Gleason Scores from Whole Slide Images [12].

El documento [12] centra su atención en el diseño de un algoritmo inteligente para la clasificación de Gleason del cáncer de próstata, a partir de imágenes microscópicas que permiten observar la estructura celular del tejido prostático. Cabe resaltar que estas imágenes del tejido prostático están teñidas de Hematoxilina-Eosina (H&E), permitiendo identificar la composición del tejido ya que tiñe en tonos azules las estructuras ácidas, y tiñe en púrpura los núcleos celulares. De esta manera, Hongming para llevar a cabo la clasificación, dispone de 312 diapositivas de patología de pacientes otorgados por la organización The Cancer Genome Atlas (TCGA). Cabe señalar que cada diapositiva de patología completa tiene un tamaño de alrededor de 1 GB y una resolución de 20x, por lo tanto, fue necesario disminuir la resolución a una de 2.5x.

Por otro lado, también se resalta en [12] que para la extracción de las características de cada imagen hace uso de descriptores de patrones binario locales estadísticos como es el CSLBP, el cual se encarga de codificar los patrones de las imágenes locales mediante el análisis de

la diferencia entre cada píxel central y los píxeles de sus vecindarios circundantes. De esta manera, al extraer las características de textura CSLBP, luego estas son integradas y utilizadas en el algoritmo de clasificación SVM, el cual se encarga de asignar diferentes puntajes de Gleason por cada diapositiva de los pacientes evaluados. Finalmente, también se destaca que el algoritmo de clasificación SVM diseñado se comparó con otros modelos de características como son: Análisis Fractal, Histograma de Haralick y Gabor. A su vez, también se comparó con máquinas VGG16-TL1, VGG16-TL2, VGG16-DL. En donde, el algoritmo de clasificación con mejor desempeño fue el SVM fue al evaluar las características de textura CSLBP, teniendo una precisión del 77.12 %

Multi-Features Classification of Prostate Carcinoma Observed in Histological Sections: Analysis of Wavelet-Based Texture and Colour Features [13].

Otro enfoque realizado [13] fue la extracción de características de textura, esto a través de convertir la imagen del modelo RGB a escala de grises y aplicar una descomposición 2D por medio la transformada Wavelet Haar. Este procedimiento consiste en tomar imágenes de biopsias del tejido prostático de tamaño 256x256 y hacer una descomposición de dicha imagen en sub-imágenes de 128x128 y a partir de estas dos descomposiciones se extraen las características de textura, tanto para la primera y segunda descomposición las características fueron las siguientes: **contraste, homogeneidad, correlación, energía, entropía, disimilitud.**

Se realizó diferentes clasificaciones binarias a través de una red neuronal artificial MLP, los grupos de clasificación se organizaron así: **Benigno vs Maligno, Benigno vs Grado 3, Benigno vs Grado 4, Benigno vs Grado 4,5, Benigno vs Grado 5.** Los resultados fueron los siguientes:

Tabla 2-1. Resultado de clasificación transformada Wavelet nivel 1 [13]

Grupos	Exactitud	Sensibilidad	Especificidad	F1 score	MCC (%)
Benigno vs Maligno	95.0	96.5	93.5	94.9	90.0
Benigno vs Grado 3	88.3	82.9	96.0	89.2	77.7
Benigno vs Grado 4	95.0	93.5	96.5	95.1	90.0
Benigno vs Grado 5	98.3	100.0	96.8	98.3	96.7
Benigno vs Grado 4/5	85.0	95.4	76.9	80.8	69.5

Tabla 2-2. Resultado de clasificación transformada Wavelet nivel 2 [13]

Grupos	Exactitud	Sensibilidad	Especificidad	F1 score	MCC (%)
Benigno vs Maligno	91.7	93.1	90.3	91.5	83.4
Benigno vs Grado 3	86.7	80.6	95.8	87.9	74.8
Benigno vs Grado 4	91.7	90.3	93.1	91.8	83.4
Benigno vs Grado 5	96.7	96.7	96.7	96.7	93.3
Benigno vs Grado 4/5	83.3	95.4	76.3	80.8	69.2

En los resultados que se registran en las Tablas 2-1 y 2.2, sobre el trabajo hecho en [13], se observa el porcentaje de acierto del sistema CAD, respecto a los dos niveles de procesamiento por medio de la transformada wavelet y la extracción de las características a través de la manipulación del modelo de color de las imágenes.

Quantitative analysis of benign and malignant tumors in histopathology: Predicting prostate cancer grading using SVM [14].

En [14] la clasificación automática de los grados Gleason se da a través de diferentes características extraídas por medio de la aplicación de procesar las imágenes de biopsias de dos maneras.

Primero se aplica el algoritmo de “K - means clustering”, que es un método de agrupamiento no supervisado que busca separar un conjunto de datos en K grupos (cluster) basados en su similitud. El objetivo es que los datos en un cluster sean similares entre sí y diferentes a los datos en otros clusters, en donde se busca separar diferentes regiones de interés, en [13] obtienen tres grupos, imágenes donde se resaltan los núcleos de las células, los lúmenes y el stroma, para la extracción de características trabajan con las imágenes donde se resaltan los núcleos de las células y lúmenes. Posterior a esto se aplica la segmentación de watershed solo al cluster donde se enfocan las células y se extraen características. Finalmente la clasificación se hace con una máquina de aprendizaje supervisado denominada SVM (Support Vector Machine), donde la clasificación se hizo de manera binaria a los siguientes grupos o clases: Benigno vs Maligno, Grado 3 vs Grado 3,4 y Grado 4 vs Grado 5, en donde se obtuvieron los siguientes resultados:

Tabla 2-3. Resultado de desempeño en [14]

Grupos	Exactitud	Sensibilidad	Especificidad	MCC(%)
Maligno vs Benigno	88.7	91.8	86.0	70.2
Grado 3 vs Grado 4,5	85.0	81.8	88.8	70.3
Grado 4 vs Grado 5	92.5	94.7	95.0	85.1

Lumen-Nuclei Ensemble Machine Learning System for Diagnosing Prostate Cancer in Histopathology Images [15].

En [15] el trabajo que hacen para llevar a cabo la clasificación es crear una arquitectura que primero separa las imágenes histopatológicas de las biopsias en 3, el parche extraído de la imagen total original, una imagen donde se enfocan en el lumen del tejido y otra en donde se resalta los núcleos en el tejido. A las 3 imágenes le extraen características diferentes las cuales son Haralick, HOG y run-length matrix respectivamente. Además, la imagen a color le extraen las características de Haralick en 3 canales después de haber cambiado su modelo de color de RGB a HSV.

Para la clasificación utilizaron una red de expertos de dos clasificadores SVM, en donde el primero fue entrenado para la clasificación con base en las características de textura del lumen y el segundo en función de los núcleos contenidos en el tejido. Cabe destacar que su enfoque fue la clasificación binaria entre grado Gleason 3 vs 4 pues son los grados según argumentan en su artículo los más difíciles de diferenciar, obteniendo así un resultado individual de 86.6% para el clasificador que fue entrenado con las características extraídas del lumen del tejido y 87.7% para el clasificador que se entrenó con las características extraídas de los núcleos presentes en el tejido. La clasificación por medio de la red de expertos superó estos resultados con un 88.9%.

Application of Deep Learning Techniques for Prostate Cancer Grading Using Histopathological Images [16]

El artículo [16] se centra en el desarrollo de máquinas profundas encargadas de detectar la agresividad del cáncer de próstata mediante el uso de imágenes de biopsias, las cuales fueron proporcionadas por los Centros Médicos de Karolinska y Radboud. Cabe destacar que utilizaron el mismo conjunto de datos que en el artículo [11], resaltando que constaba de 10616 imágenes, cada una clasificada con un determinado grado de severidad del cáncer, representado en una escala de 0 a 5. A su vez, en este artículo también resalta que es necesario hacer un reescalamiento de todas las imágenes, mencionando que todas las imágenes se reescalaron a una dimensión de 512x512 píxeles. Por otra parte, en este artículo se mencionan que utilizaron diferentes modelos pre entrenados, a través de Transfer Learning, en máquinas como MobileNetV2, ResNet50, DenseNet121, DenseNet169, VGG-16, VGG-19, Xception, InceptionV3, InceptionResNetV2 y EfficientNetB7, en donde obtuvieron una exactitud de 62.75%, 67.16%, 72.86%, 76.44%, 81.57%, 84.64%, 88.31%, 88.46%, 88.97%, 90.90%, respectivamente.

2.1.3. Discusión

Teniendo en cuenta los antecedentes consultados se encontró que antes de diseñar un algoritmo de clasificación, es necesario tener claro qué técnicas de preprocesamiento y herramientas permiten extraer las características de cada imagen de forma adecuada y además a eso tener opciones pues son muchas las configuraciones que se pueden hacer y formas de proceder. De esta manera, ya conociendo las características de cada imagen facilita la detección de cáncer de próstata debido a que se pueden encontrar patrones en el tejido prostático, conociendo como es un tejido sano y un tejido maligno, así como también conociendo las características de una imagen histopatología como características en función de la textura, el color, estadísticas por canal R, G y B, morfológicas según la forma de las glándulas y núcleos.

A su vez, se evidencia la importancia del uso de algoritmos de Deep Learning y Redes Neuronales debido a que permite automatizar tareas, las cuales, para el caso de detección y clasificación por el grado Gleason, estas tareas son ejercidas por un especialista, el cual tiene por trabajo analizar cada imagen. De tal manera, que, si un especialista tiene como tarea clasificar 10 muestras de pacientes, le tomará una gran cantidad de tiempo para analizar cada muestra de paciente comparado a si una máquina es capaz de detectar y clasificar cada imagen, gracias a la capacidad de procesar u ofrecer un tiempo de detección en o menos de un segundo. Además, por el lado de las máquinas superficiales, también se reconocen en estas que si se desea llevar a cabo un buen trabajo de clasificación es importante cada paso y cada acción llevada a cabo en cada etapa, pues al ser la persona que realiza cada etapa de forma manual o de forma consciente a diferencia de las redes profundas, el proceder y direccionamiento de las acciones para una extracción de características relevantes y discriminantes tiene suma importancia para que la generalización y diferenciación entre clases sean logradas a través de las máquinas superficiales.

Por otra parte, al analizar cada artículo, se evidencia que la investigación [10] obtuvo un algoritmo con una precisión del 89.2% al trabajar con todas las imágenes enteras. En donde su algoritmo de clasificación fue una SVM junto con el clasificador RF. Para el caso de la investigación [12], el cual también implementa una estructura SVM no logró una precisión mayor 80%, alcanzando solo una precisión del 77.12 %. Estas diferencias en los resultados de clasificación aún así hayan usado el mismo clasificador en algunos casos, se da debido a diferencia en la metodología de solución aplicada, como diferente uso de base de datos, diferentes técnicas de preprocesamiento, segmentación, extracción de características de una a otra investigación, diferente sintonización de parámetros de un clasificador a otro, lo que resulta en resultados diferentes, estas diferencias se pueden ver con mejor claridad en la Tabla 2-4.

De los artículos [13] y [14], se destaca el constante uso de la descomposición de los canales R, G, B y cómo el combinar una técnica procesamiento del color de una imagen con una técnica de procesamiento de la textura de esta misma, generan resultados de desempeño por lo general arriba del 83% (solo se presentó una vez según antecedentes consultados) y

alcanzando desempeños de hasta 98.3%, con el algoritmo MLP, por medio de una clasificación binaria entre tejidos benignos versus grado 5.

Ahora bien, también es necesario resaltar que en los 2 estudios, [11] y [16], están enfocados en el desarrollo de máquinas profundas, siendo estas máquinas las encargadas de detectar la agresividad del cáncer de próstata mediante el uso de imágenes de biopsias. En ambos artículos se resalta que utilizaron el mismo dataset proporcionado por los Centros Médicos de Karolinska y Radboud, entregando un total de 10616 imágenes. Las diferencias de estos artículos se encuentran que en [11], las imágenes fueron reescaladas a una dimensión de 224x224 píxeles, en donde el mejor desempeño lo encuentra en la máquina InceptionResNet V2 con un 84.99% de exactitud. Por otro lado, para el caso del artículo [14], menciona que las imágenes del tejido completo fueron modificadas a una dimensión de 512x512 píxeles, teniendo su mejor desempeño en la máquina EfficientNetB7 con 90.90% de exactitud.

Finalmente, un consejo que se destaca en los artículos [11], [13], [14] y [16] es el determinar un tamaño de parche y ajuste en la resolución de cada imagen, debido a que la mayoría de las imágenes microscópicas del tejido prostático tiene un tamaño en giga píxeles, por tanto, afecta en el procesamiento y análisis puesto que el computador necesita una mayor capacidad de cálculo y memoria. A su vez, es importante evitar el Overfitting para el diseño del sistema, lo que puede dar lugar a un bajo rendimiento y disminuiría la precisión del sistema a la hora de clasificar las distintas imágenes, si no se tiene en cuenta.

Tabla 2-4. Resumen de los antecedentes.

Referencia	Aplicación	Tipo de imagen	Origen y calidad del dataset	Preprocesamiento	Clasificador	Precisión
[9]	Detección	MR	- Imágenes: 339 - Tomadas por: The Cancer Imaging Archive (TCIA)	- Diffusion weighted imaging (DWI) - Escalamiento	- SVM - k-NN - RaF	- SVM: 80.83% - k-NN: 76.83% - RF: 75.89%
[10]	Detección Clasificación	Microscópica teñidas de H&E	- Imágenes: 213 - Tomadas por: The Cancer Genome Atlas (TCGA).	- Descomposición en 3 canales (R,G,B) - Subsampling - Normalización del color - Reducción de resolución	- SVM - RF	-Precisión: 89.2 %
[11]	Detección Clasificación	Microscópica teñidas de H&E	- Imágenes: 10616 - Tomadas por: Centros Médicos de Karolinska y Radboud	- Escalamiento	MobileNet V2 InceptionResNet V2 DenseNet 169 ResNet101 V2 NASNetMobile	MobileNet V2: 80.18%. InceptionResNet V2: 84.99%. DenseNet 169: 83.98%. ResNet101 V2: 81.50%. NASNetMobile: 75.13%.
[12]	Detección Clasificación	Microscópica teñidas de H&E	- Imágenes: 312 - Tomadas por: The Cancer Genome Atlas (TCGA).	- Deconvolución en color - Binarización por método Otsu -Reducción de resolución	- SVM con FA - SVM con HHG - SVM con HHG+PCA - SVM con - VGG16-TL1 - VGG16-TL2 - VGG16-DL	- SVM con FA: 61.1% - HHG:63.5% - HHG+PCA: 65.1% - VGG16-TL1: 64.5% - VGG16-TL2:69.1% - VGG16-DL:65.4% - SVM:77.1 %
[13]	Clasificación binaria	Microscópica teñidas de H&E	- Imagenes: 400 - Tomadas de Severance Hospital of Yonsei University	- Descomposición en 3 canales (R,G,B) - Conversión a escala de grises y descomposición wavelet 2D	MLP	Benigno vs Maligno: 95.0 % Benigno vs Grado 3: 88.3% Benigno vs Grado 4: 95.0% Benigno vs Grado 5: 98.3% Grado 3 vs Grado 4/5: 85.0%
[14]	Clasificación binaria	Microscópica teñidas de H&E	- Imagenes: 400 - Tomadas de Severance Hospital of Yonsei University	- Segmentación por K-mean clustering - Segmentación por método de Watershed	SVM	Maligno vs Benigno: 88.7% Grado 3 vs Grado 4,5: 85.0% Grado 4 vs Grado 5: 92.5%
[15]	Clasificación binaria	Microscópicas teñidas de H&E	-Imágenes: 149 no especifica fuente del dataset	K-mean clustering	Red de expertos con SVM	88.9%
[16]	Detección Clasificación	Microscópica teñidas de H&E	- Imágenes: 10616 - Tomadas por: Centros Médicos de Karolinska y Radboud	- Escalamiento	MobileNetV2 ResNet50 DenseNet121 DenseNet169, VGG-16 VGG-19 Xception InceptionV3 InceptionResNetV2 EfficientNetB7	MobileNetV2: 62.75%. ResNet50: 67.16% DenseNet121: 72.86% DenseNet169: 76.44% VGG-16: 81.57% VGG-19: 84.64% Xception: 88.31% InceptionV3: 88.46% InceptionResNetV2: 88.97% EfficientNetB7: 90.90%

2.2. Marco teórico

El objetivo general de este trabajo es implementar un sistema de procesamiento de imágenes microscópicas en el espectro visible para detectar la presencia o ausencia y la severidad del cáncer de próstata. De esta manera, para realizar este trabajo es necesario tener en cuenta algunos conceptos generales sobre cáncer de próstata, biopsia, clasificación Gleason, preprocesamiento de imágenes, segmentación de imágenes, extracción de características y reconocimiento de patrones.

2.2.1. Cáncer de próstata

El cáncer es una enfermedad debida a la reproducción rápida y descontrolada de las células de los diferentes órganos, tejidos o demás partes funcionales y vitales que componen el cuerpo humano. El cáncer de próstata tiene su origen debido al crecimiento sin control de las células de la próstata, la cual es una glándula presente sólo en los hombres y cuya severidad del cáncer se divide en diferentes grados [17].

2.2.1.1. Adenocarcinoma

El adenocarcinoma de próstata es el tipo de cáncer más común entre los hombres y generalmente se desarrolla lentamente, a menudo sin síntomas, en las etapas iniciales, puede crecer y propagarse con el tiempo a otras partes del cuerpo, especialmente en los ganglios linfáticos y los huesos.[18]

El diagnóstico se realiza por lo general mediante pruebas de detección, como el examen de antígeno prostático específico (PSA) y el examen rectal digital (DRE), seguido de pruebas de confirmación, como la biopsia de próstata.

Dependiendo de la agresividad de este tipo de cáncer puede variar el tratamiento, el cual puede incluir vigilancia activa, cirugía, radioterapia, terapia hormonal, quimioterapia o una combinación de estos. [19]

2.2.1.2. Biopsia

Una biopsia es una extracción de células o tejidos para ser examinados por un patólogo. Este estudio puede ser a través de un microscopio o puede que se realice demás pruebas con las células y/o tejido extraído.

Existen diferentes tipos de biopsias:

- **Biopsia por incisión:** Consiste en la extracción de solo una muestra del tejido.
- **Biopsia por escisión:** Extracción de un área o masa de la cual se tienen sospechas de presencia de cáncer.
- **Biopsia con aguja:** Se extrae una muestra de tejido o líquido por medio de una aguja. La aguja puede ser ancha, en este caso el proceso lleva el nombre de biopsia por punción con aguja ancha. Otro posible nombre es la biopsia por aspiración con aguja fina cuando la aguja es precisamente fina [20].

2.2.1.3. Coloración H&E

La tinción de hematoxilina (H) y eosina (E) o H&E como se le conoce es una técnica de tinción histológica comúnmente utilizada en patología con el fin de resaltar diferentes componentes de los tejidos en muestras histopatológicas por medio de estos dos colorantes, hematoxilina y eosina.

Esta tinción permite a los patólogos evaluar e identificar diferentes características histológicas, como la arquitectura celular, presencia de células anormales, inflamación y demás cambios a nivel microscópicos. [21]

2.2.1.4. Clasificación Gleason

El cáncer de próstata una vez es detectado se debe diagnosticar la severidad de este, pues según la severidad varía el tratamiento. Esto se lleva a cabo mediante la asignación de un grado en función del patrón de crecimiento del tejido afectado; teniendo desde el grado uno hasta el cinco, donde el grado 1 es el mínimo y 5 el máximo [22]. Debido a que los patrones de crecimiento que se presentan en el tejido son más de uno, la clasificación se hace tomando en cuenta el patrón de tejido mayoritario y minoritario, lo que da un nuevo puntaje de Gleason y posteriormente se clasifica dentro de los 5 grados posibles, como se ilustra con la siguiente imagen:

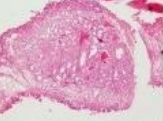
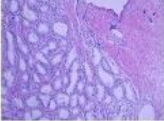
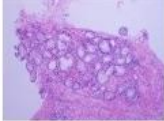
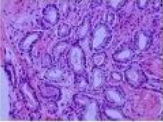
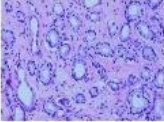
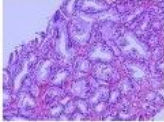
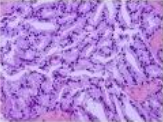
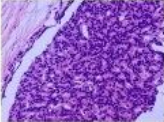
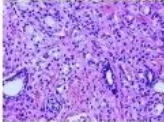
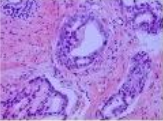
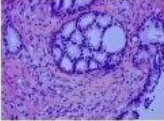
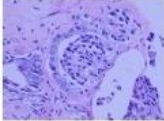
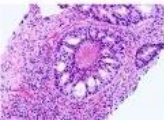
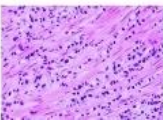
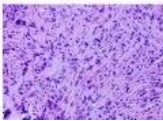

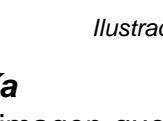
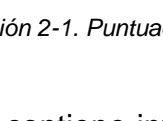
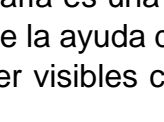
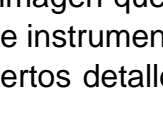
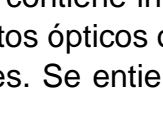



			Puntuación del patrón de Gleason 1-3 glándulas individuales distintas y discretas	Puntuación de Gleason del tejido ≤ 6	Grupo de Grado I
					
			Puntuación del patrón de Gleason 4 glándulas fusionadas, cribriformes o mal formadas	Puntuación de Gleason del tejido 3 + 4 = 7	Grupo de Grado II
					
				Puntuación de Gleason del tejido 4 + 4 = 8 3 + 5 = 8 5 + 3 = 8	Grupo de Grado III
					
			Puntuación del patrón de Gleason 5 necrosis comedoniaria, cordones, láminas, redes sólidas y células individuales	Puntuación de Gleason del tejido 4 + 5 = 9 5 + 4 = 9 5 + 5 = 10	Grupo de Grado IV
					

Ilustración 2-1. Puntuación Gleason [23]

2.2.1.5. Micrografía

Una micrografía es una imagen que contiene información que no es visible a simple vista y que mediante la ayuda de instrumentos ópticos o electrónicos como lupas y microscopios se pueden hacer visibles ciertos detalles. Se entiende entonces por fotografía microscópica o

fotomicrografía al conjunto de técnicas que permiten obtener estas imágenes con una ampliación mínima de diez veces [24].

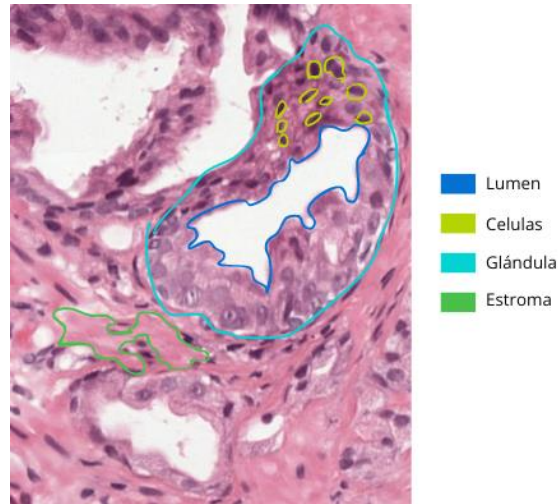


Ilustración 2-2. Partes que se pueden encontrar en una imagen de biopsia de próstata

En la Ilustración 2-2, se evidencia las diferentes partes y componentes que se encuentran en una biopsia de próstata, siendo estas:

- **Células:** Son principalmente células epiteliales que forman las glándulas prostáticas. Estas células están especializadas en la producción y secreción de PSA, proteína que se produce en la próstata. [25]
- **Lumen:** Es el espacio hueco presente en una estructura tubular, como lo son los conductos glandulares de la próstata. El lumen se visualiza como un área clara dentro del tejido, cuando es observada bajo un microscopio, rodeado por las células epiteliales, el tamaño varía en función del estado de la glándula prostática. [25]
- **Glándula:** Están compuestas por células epiteliales, son estructuras alargadas o redondeadas, rodeadas por tejido conectivo, que contienen un lumen central. Puede variar en tamaño y forma dependiendo de la edad, estado de salud de la persona. [25]
- **Estroma:** El estroma es el tejido conectivo que proporciona soporte estructural a los celulares glandulares y los demás elementos presentes en el tejido prostático. Se ubica entre las glándulas y alrededor de ellas, apareciendo como tejido fibroso con células dispersas. [25]

2.2.2. Preprocesamiento de imágenes

El preprocesamiento de imágenes es una de las primeras etapas en las máquinas superficiales. Este es el proceso de modificar un conjunto de imágenes durante el procesamiento para que la información que contienen estas imágenes sea más clara y fácil de interpretar, analizar y utilizar en las etapas posteriores.

Luego, en esta etapa, encontrará trabajos que utilizan enfoques de reducción de ruido, nitidez, contraste, mejora de la opacidad, mejora de la detección de bordes y otras técnicas, según el tipo de imagen y lo que el investigador desee hacer.

En el caso de las máquinas profundas, el preprocesamiento es menos extenso, ya que es la máquina la que extrae la información que se considera relevante para realizar la tarea de clasificación. Se puede encontrar como preprocesamiento en las máquinas profundas la conversión de color a escala de grises y la normalización de imágenes [26].

Para este trabajo se tuvo en cuenta dos formas de aplicar el preprocesamiento a las imágenes, en escala de grises que consiste en hacer manipulaciones a las imágenes para realzar características en la imagen, mientras se mantiene la imagen en un solo plano, el plano de escala de grises. Por otro lado, el preprocesamiento con imágenes a color, implica también manipulaciones a la imagen pero teniendo en cuenta los 3 canales R, G y B.

2.2.2.1. Preprocesamiento en escala de grises

2.2.2.1.1. Corrección Gamma

La corrección gamma es una operación no lineal que se utiliza para codificar y decodificar valores de luminancia o color en sistemas de imágenes fijas o de vídeo. Este proceso se basa en la forma no lineal en que los humanos perciben la luz y el color. Esto se utiliza para optimizar el uso de bits al codificar imágenes y el ancho de banda utilizado para la transferencia. Las expresiones de la ley de potencia no lineal se aplican para codificar o decodificar valores de luminancia o color. Hay muchas formas de realizar la corrección gamma, incluida la manipulación de hasta cuatro elementos para lograr la codificación gamma. [27].

2.2.2.1.2. Transformada Wavelet multi nivel

Las transformadas wavelets multinivel son poderosas herramientas de análisis que descomponen señales o imágenes en componentes de diferentes escalas o resoluciones. Esto implica aplicar secuencialmente filtros de paso bajo y paso alto y submuestreo para descomponer la imagen en un conjunto de coeficientes wavelet. Estos coeficientes representan la imagen original en diferentes niveles de detalle, desde el más general (aproximación de bajo nivel) hasta el más específico (detalle de alto nivel). En cada nivel de esta descomposición, la imagen se divide en una serie de bandas o sub-bandas, cada una de las cuales captura características específicas, como bordes horizontales, verticales o diagonales.

En el contexto de la extracción de características de imágenes, las transformadas wavelets multinivel son particularmente útiles porque pueden separar características en diferentes escalas y orientaciones, haciéndolas resistentes al ruido y las variaciones de iluminación. Los coeficientes wavelet aprovechan el hecho de que, si bien las aproximaciones de bajo nivel capturan la esencia de una imagen, los detalles de alto nivel pueden contener información redundante o menos importante. Se pueden utilizar directamente como una característica para la clasificación de imágenes, el análisis de texturas o la compresión de imágenes. Esto permite reconstruir imágenes de diferentes niveles de calidad y eliminar coeficientes correspondientes a pequeños detalles. Esta es una técnica común para la compresión de imágenes que reduce el tamaño del archivo y al mismo tiempo preserva la funcionalidad clave. [28].

2.2.2.2. Preprocesamiento a color

2.2.2.2.1. Normalización de la tinción

La normalización de la tinción en imágenes médicas o histopatológicas, especialmente aquellas que son teñidas por medio de colorantes que se aplican para resaltar zonas de interés para analizar, es un paso crítico en el preprocesamiento de imágenes. Este proceso centra su atención en corregir las variaciones en intensidad y color de las imágenes debido a las tinciones aplicadas al tejido. Esto no solo mejora la visualización de la imagen, sino optimiza la precisión de los algoritmos que se usan posterior a esta etapa en donde su foco es la extracción de características relevantes del problema de clasificación. [29]

2.2.3. Segmentación de imágenes

La segmentación de imágenes, al igual que el preprocesamiento de imágenes, es un paso que se puede considerar cuando se utilizan máquinas de aprendizaje superficial. Una vez que se realiza el preprocesamiento de la imagen y se mejora y modifica la imagen para resaltar la información de interés, se resaltan aún más las regiones o regiones de interés y se utilizan diversas técnicas para segmentar la imagen, que tiene un valor aún mayor cuando se usa.

Esta es una fase continua de trabajo para mejorar aún más las imágenes de modo que estas técnicas puedan distinguir mejor entre la información de interés de aquella que no es relevante para la clasificación deseada. [30].

2.2.3.1. Segmentación método Otsu

En el procesamiento de imágenes, la segmentación por umbralización de Otsu divide automáticamente los píxeles de una imagen en dos grupos. Utiliza el histograma de la imagen para encontrar el umbral ideal que minimice la variabilidad dentro de cada grupo y maximice la separación entre los grupos. Se puede dividir automática y precisamente la imagen en áreas de interés y fondo al encontrar este umbral [31].

2.2.3.2. Segmentación por contornos activos

En el procesamiento de imágenes, la segmentación por contornos activos delinea y extrae automáticamente los bordes de los objetos relevantes de una imagen. Consiste en inicializar un contorno cercano al objeto deseado y ajustarlo iterativamente hacia los bordes exactos del objeto. Esto se logra mediante la minimización de una función de energía que combina factores relacionados con la suavidad del contorno y su proximidad a los bordes. Los bordes activos se pueden ajustar a cambios de forma, bordes discontinuos o ruidosos y se pueden controlar por datos de características de imagen, como gradientes o regiones de interés. Aplicaciones como la segmentación de imágenes médicas, la detección de objetos y el seguimiento de objetos en movimiento incluyen esta técnica [32].

2.2.3.3. Segmentación por método Watershed

El método de Watershed es una técnica de procesamiento de imágenes utilizada para la segmentación y detección de bordes, que se basa en la simulación del proceso de inundación de una cuenca hidrográfica. En este método, se identifican los mínimos locales en la imagen

y se simula el llenado de estas cuencas con agua, lo que permite unir las cuencas en los puntos más bajos y formar límites precisos entre ellas. Sin embargo, este método puede generar segmentaciones excesivas si no se aplica adecuadamente [33].

2.2.3.4. Segmentación basada en agrupamiento – clustering

La segmentación basada en clustering es una técnica de procesamiento de imágenes que agrupa píxeles similares en la misma clase o clúster en función de características comunes como la intensidad del color o la textura. El objetivo es dividir la imagen en segmentos de modo que los píxeles de un solo grupo sean más similares entre sí que los píxeles de otros grupos. Esto se logra mediante algoritmos de agrupamiento que organizan los datos en grupos en función de la distancia o similitud entre elementos. Uno de los algoritmos de agrupamiento más utilizados para la segmentación de imágenes es K-means.

Este método intenta dividir la imagen en k grupos. Cada píxel pertenece al grupo con la media más cercana (centroide del grupo). Este proceso se repite actualizando el centroide de cada grupo hasta que la asignación de píxeles al grupo no cambia o hasta que se cumple algún otro criterio de terminación [34].

2.2.3.5. Métricas de evaluación

Cuando se segmentan imágenes, siempre es importante contar con una imagen patrón o guía, en inglés Ground Truth. Pues de esta forma es posible llevar una guía de la calidad de la segmentación aplicada por el método que se esté llevando a cabo este proceso. Para este trabajo se aplicó a las segmentaciones llevadas a cabo los siguientes índices con el fin de garantizar una representación lo menos alejada a las imágenes originales [35]:

2.2.3.5.1. Índice de Jaccard

El índice de Jaccard es una métrica comúnmente utilizada para evaluar la similitud entre dos conjuntos, produce un valor en el rango de 0 a 1, donde 0 indica una completa disimilitud entre los conjuntos y 1 similitud perfecta. Este índice es útil para comparar similitud entre dos conjuntos de elementos como también la superposición entre áreas delimitadas en imágenes. [35]

2.2.3.5.2. Índice de Dice

El índice de Dice calcula la disimilitud o similitud entre imágenes, útil en aplicaciones donde hay desequilibrio entre el tamaño de los conjuntos, debido a que penaliza menos los falsos negativos en comparación que el índice de Jaccard. [35]

2.2.3.5.3. Índice BF score

El índice BF Score arroja resultados de la segmentación respecto a la precisión de la delimitación de los contornos de los objetos, en donde combina la precisión y exhaustividad de los contornos predichos en comparación al Ground Truth. [35]

2.2.4. Extracción de características

La fase de extracción de características ocurre en el flujo de trabajo de una máquina superficial después de que la imagen haya sido preprocesada y segmentada, o simplemente

después de haber sido preprocesada o segmentada, dependiendo de cómo se creó la arquitectura del sistema de clasificación. En esta fase se obtiene un vector $1 \times n$. Aquí, n es el conjunto correspondiente de valores almacenados que se utilizan para representar la información contenida en cada imagen a clasificar. Estas características pueden ser de diferentes tipos. Cuando se trabaja con imágenes, las funciones estadísticas y de textura se utilizan con mayor frecuencia [36].

La idea de la extracción de características es una vez preprocesada la imagen, mejorada y resaltada la región o zonas de interés se extrae la información más relevante y representativa por las clases que se tengan en el problema de clasificación para posteriormente entrenar a los clasificadores a usar y así crear el espacio de características y a su vez la función discriminante que permite hacer la respectiva separación de cada clase para una correcta clasificación.

2.2.4.1. Características por contorno

Las características de contorno son atributos descriptores de propiedades de los bordes o límites de un objeto o superficie. Estas características son usadas para describir la forma y estructura de los objetos en una imagen o superficie. Estas características pueden incluir información sobre la longitud total del contorno, cantidad de curvatura, relación entre eje mayor y menor, concavidad de las áreas de contorno, entre otras.

Estas características son importantes en análisis de imágenes para tareas de reconocimiento de patrones, detección de objetos, inspección de calidad y demás procesos de visión por computadora.

2.2.4.2. Características por textura y momentos estadísticos

Estas características son útiles para cuantificar y caracterizar la variabilidad en la apariencia de una superficie. Son fundamentales en aplicaciones donde la apariencia superficial es un factor crítico, como lo es la apariencia de tejidos en imágenes médicas. Estas características pueden capturar y describir información sobre rugosidad, suavidad, regularidad, granularidad, orientación, que a través de su análisis las máquinas de aprendizaje automático pueden detectar diferencias entre clases y llegar a clasificaciones concretas.

Para este trabajo se extrajeron 2 tipos de características capaces de describir los datos a través de textura, por medio del color y en escala de grises.

2.2.4.2.1. Matriz GLCM

La matriz GLCM es una herramienta estadística para examinar la textura de imágenes. La textura se puede describir como un patrón de intensidad que se repite dentro de una imagen y es una característica importante en muchas aplicaciones de procesamiento de imágenes y visión por computadora. Un GLCM es una Tabla que cuantifica con qué frecuencia aparecen diferentes combinaciones de niveles de intensidad de píxeles (o niveles de gris) en una imagen, organizadas en una relación espacial particular.

Esta matriz se construye seleccionando un par de píxeles a la vez, considerando uno como referencia y el otro a una distancia ($d = 1$) y ángulos específicos ($0^\circ, 45^\circ, 90^\circ, 135^\circ$). Luego se cuenta cuántas veces cada combinación de niveles de gris ocurre en esta disposición. Estas métricas proporcionan información cuantitativa sobre la textura de la imagen que puede ser útil para la clasificación, el análisis de patrones y el diagnóstico médico por imagen, entre otros [37].

2.2.4.2.2. Estadísticas por canal R, G y B

Este tipo de extracción implica analizar la distribución y variación de los píxeles en cada componente de color R, G y B en una imagen. La técnica se basa en que es posible capturar diferentes estructuras y texturas en una imagen a través de su análisis por canal que, analizando por completo la imagen, detalles que pueden aportar a características con poder descriptivo para el problema de clasificación. Al calcular estadísticas como media, desviación estándar, asimetría y curtosis para cada canal de color, es posible obtener características que capturan tanto información de color como textura de la imagen, permitiendo una caracterización detallada de las regiones y estructuras presentes en la imagen a través de los 3 aportes que pueden llegar a brindar los 3 canales R, G y B. [38]

2.2.5. Reconocimiento de patrones

El reconocimiento de patrones es una disciplina científica que tiene por objetivo la clasificación de objetos o medidas lo que se conoce como patrones. Estos patrones son extraídos de imágenes y/o señales de audio en diferentes categorías o clases según cómo se busque aplicar la técnica, el objetivo final es llegar a una distinción entre una misma medida, ya sea imagen, señales de todo tipo y demás a través de unas características ya sean definidas o encontradas por el propio sistema [39].

2.2.5.1. Características

En reconocimiento de patrones una característica es un parámetro o indicador que puede ser conocido o no conocido dependiendo del tipo de aprendizaje, que permite la clasificación de los objetos o patrones y que puede llegar a representar fielmente la o las diferentes clases que son de interés a clasificar [40].

2.2.5.2. Tipos de aprendizaje

Los tipos de aprendizaje difieren dependiendo si se conoce o no las características que usa el clasificador para reconocer los patrones, estos se dividen en aprendizaje supervisado que es la clasificación de los patrones por medio del conocimiento previo de las características y la clase a la que pertenecen, es decir, el clasificador se entrena con un conjunto de datos de entrada con salidas conocidas.

El aprendizaje no supervisado es la clasificación de los patrones teniendo las características, pero no la clase a la que pertenecen, la clase es el tipo de patrón que es cada imagen dependiendo de las características. En este tipo de aprendizaje el conjunto de entrenamiento no dispone de salidas conocidas.

En el aprendizaje semi supervisado se tiene tanto características conocidas como desconocidas, es decir, tiene entradas con su respectiva salida, pero entrada de las cuales no tiene la salida esperada a priori [41].

2.2.6. Máquinas de aprendizaje superficial

Las máquinas de aprendizaje superficial son modelos de aprendizaje automático que aprenden a través de la extracción manual de características relevantes de los datos de entrada. Estas características son trabajo del especialista, ingeniero, científico de datos y es quién se encarga de que este conjunto de descriptores, sean lo más apropiado posible para que la máquina tenga una buena generalización en su aprendizaje y de forma implícita posea el menor sesgo o bias posible al momento de clasificar los datos de entrada.

Para este trabajo se optó por las siguientes máquinas de aprendizaje superficial:

2.2.6.1. Naive Bayes

Naive Bayes (NB), es un clasificador que basa su función discriminante o su regla de decisión con base en funciones de densidad estadísticas. Basa su regla decisión acorde al teorema de Bayes. A pesar de que es un clasificador que de manera óptima responde muy bien a datos que sigan una distribución normal conocida y se pueda aproximar por medio de alguna función de densidad; aporta resultados apropiados para datos que no sigan o se puedan modelar bajo alguna distribución de probabilidad, pues esta versión ingenua asume independencia de clases e ignora si estas siguen o no una distribución de probabilidad, por lo que es un acercamiento al problema de clasificación de manera estadística considerado a probar [42].

2.2.6.2. k-Nearest Neighbors

k-Nearest Neighbors (kNN), basa su funcionamiento en medir qué tan cercano es un nuevo dato respecto a las demás clases, el valor de k es un hiperparámetro importante en este tipo de clasificadores pues se debe tener en cuenta que se haga una medida de la cercanía óptima en función de cómo se calculan las distancias, como también que no exceda los límites de datos por clases [43].

2.2.6.3. Support Vector Machine

Support Vector Machine (SVM), es un algoritmo utilizado para clasificar datos. Su objetivo es encontrar un hiperplano que separe óptimamente las clases en un espacio de alta dimensión. Para ello, busca maximizar la distancia entre los puntos de datos más cercanos de diferentes clases, conocidos como vectores de soporte. Dependiendo de la naturaleza del problema, puede utilizar diferentes funciones de kernel para mapear los datos a un espacio de mayor dimensionalidad. Una vez encontrado el hiperplano óptimo, SVM puede clasificar nuevos datos según su posición relativa a este hiperplano. Es un método efectivo para resolver problemas de clasificación lineal y no lineal [44].

2.2.6.4. Multilayer Perceptron

Multilayer Perceptron (MLP), es un tipo de red neuronal con al menos tres capas: entrada, una o más ocultas y salida. Cada capa tiene nodos con conexiones ponderadas. La

información fluye de la entrada a través de las capas ocultas hacia la salida, y las conexiones se ajustan durante el entrenamiento para minimizar las diferencias entre las salidas predichas y las deseadas. Esto permite que la red aprenda representaciones complejas y no lineales de los datos, lo cual es una opción para contemplar, pues puede que la solución implique un tratamiento no lineal, en vez de lineal como podría darse a través de Bayes o kNN [45].

2.2.7. Máquinas de aprendizaje profundo

Las máquinas de aprendizaje profundo son también modelos de aprendizaje automático cuyo enfoque estructural es diferente al de las máquinas superficiales, pues hacen el tratamiento de los datos por capas cada una constituida por capas convolucionales y una cantidad de neuronas determinadas, las cuales tienen la capacidad de aprender de manera autónoma los datos descriptivos o características de las capas convolucionales que a diferencia de las máquinas de aprendizaje superficial es el especialista el encargado de extraer las características.

Para este trabajo se optó por las siguientes máquinas de aprendizaje profundo:

2.2.7.1. AlexNet

AlexNet es una arquitectura de red neuronal convolucional (CNN) que ganó fama después de ganar la competencia ImageNet Large Scale Visual Recognition Challenge en 2012. Diseñada por Alex Krizhevsky, Ilya Sutskever y Geoffrey Hinton, AlexNet es significativamente más profunda y grande que las CNNs previas a ella.

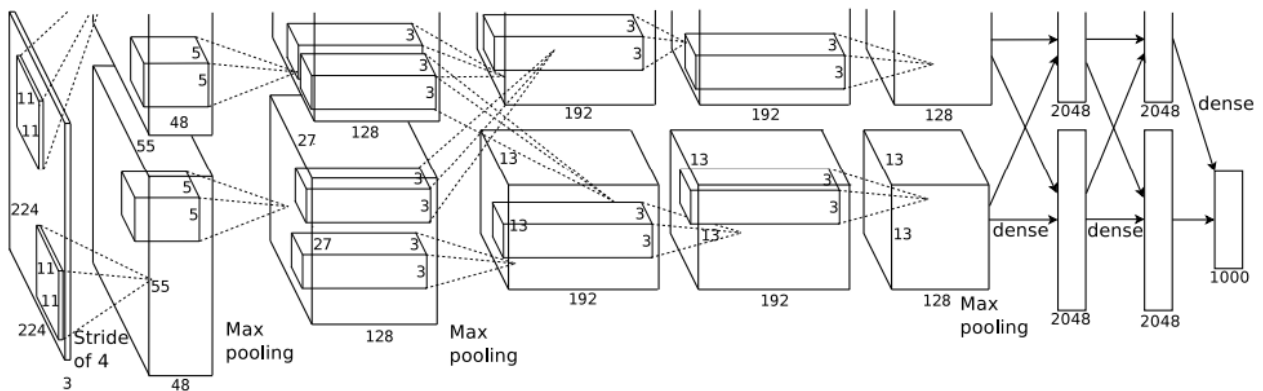


Ilustración 2-3. Arquitectura red AlexNet tomada de [46]

Fue una de las primeras en utilizar unidades de rectificación lineal (ReLU) para las activaciones, además de técnicas como el "dropout" para reducir el sobreajuste [46].

2.2.7.2. ResNet-50

ResNet-50 es una versión del modelo Residual Network (ResNet) que fue introducido por Kaiming He et al. en 2015. Tiene 50 capas de profundidad y es conocido por sus "conexiones residuales".

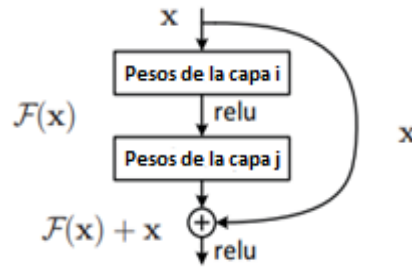


Ilustración 2-4. Ejemplo de conexión residual, tomado de [47]

Estas conexiones permiten señales de entrada saltarse una o más capas, lo que ayuda a aliviar el problema de desvanecimiento de gradiente y permite entrenar redes mucho más profundas. Estas conexiones actúan como atajos que permiten que el gradiente fluya a través de la red sin degradarse [47].

2.2.7.3. EfficientNetB0

EfficientNetB0 es la versión base de la familia de modelos EfficientNet, propuesta por Mingxing Tan y Quoc V. Le en 2019. Los modelos EfficientNet utilizan una metodología sistemática para escalar las redes en tres dimensiones: anchura, profundidad y resolución de imagen.

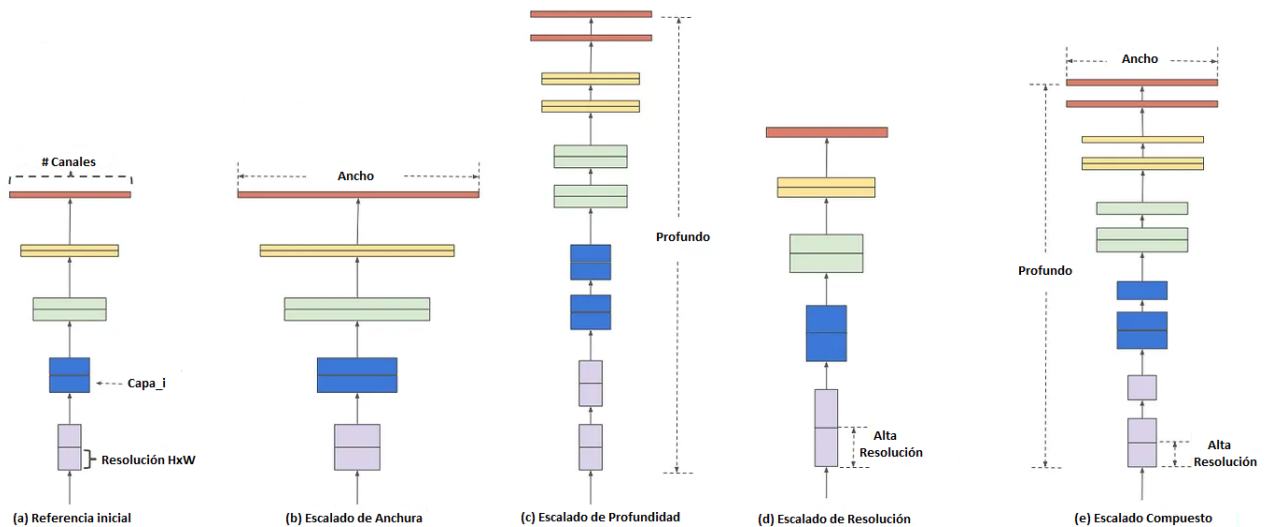


Ilustración 2-5. a) Modelo de CNN convencional, b) – d) Modelos que solo tratan una dimensión en las imágenes, e) Modelo propuesto en [48].

Estos modelos son conocidos por ser altamente eficientes, logrando un excelente rendimiento con significativamente menos parámetros y FLOPS (operaciones de punto flotante por segundo) en comparación con otras arquitecturas [48].

2.2.8. Métricas de evaluación

Las métricas de evaluación son una medida de complemento que representa y da fiabilidad sobre el desempeño de la clasificación hecha por una máquina superficial o profunda, ya sea clasificando las clases correctas o clases negativas como también la relación entre estas que en conjunto dan una perspectiva más completa del comportamiento y desempeño de las máquinas [49].

2.2.8.1. Precisión

La precisión mide la proporción de identificaciones positivas que fueron realmente correctas. Se calcula como el número de verdaderos positivos dividido por la suma de verdaderos positivos y falsos positivos. [49]

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (2-1)$$

2.2.8.2. Sensibilidad

La sensibilidad mide la capacidad del modelo para identificar correctamente los casos positivos de entre todos los casos realmente positivos. Se calcula como el número de verdaderos positivos dividido por la suma de verdaderos positivos y falsos negativos. [49]

$$\text{Sensibilidad} = \frac{TP}{TP + FN} \quad (2-2)$$

2.2.8.3. F1 score

El F1 Score es una medida de la precisión de una prueba. Considera tanto la precisión como la sensibilidad del modelo para calcular la puntuación. Es la media armónica de la precisión y la sensibilidad. [49]

$$F1 \text{ score} = 2 * \frac{\text{Sensibilidad} * \text{Precisión}}{\text{Sensibilidad} + \text{Precisión}} \quad (2-3)$$

2.2.8.4. Exactitud

La exactitud mide la proporción de predicciones correctas (tanto positivas como negativas) entre el total de casos. Se calcula como la suma de verdaderos positivos y verdaderos negativos dividida por el total de casos. [49]

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-4)$$

2.2.9. Discusión

Los conceptos discutidos en esta sección son importantes pues representan la base o conocimiento inicial necesario para abordar el problema de clasificación en relación con las técnicas de visión artificial y reconocimiento de patrones que se podrían llegar a utilizar en la solución del problema, así como también la funcionalidad y alcance de los diferentes clasificadores considerados a usar al igual que la forma en cómo se verificará el desempeño de los desempeños alcanzados.

Las imágenes disponibles y que comúnmente son usadas en este tipo de problemas y en la generación de sistemas CAD son importantes pues son la fuente de información de análisis y de donde se extraen relaciones que resultan relevantes para la clasificación, de donde se desprenden diferentes tareas a realizar, desde el preprocesamiento, segmentación, extracción de características hasta la clasificación, cada etapa con técnicas, que aplicadas según se requiera y, que en conjunto generan un sistema capaz de llevar a cabo la tarea de clasificación, en este caso del grado Gleason.

Un aspecto también a tomar en cuenta es el funcionamiento de la asignación del grado Gleason, pues su comprensión permitirá el cómo se hace la extracción de características en el caso de las máquinas superficiales para así distinguir un grado del otro, pues hay criterios para cada grado Gleason, que es un aspecto clave al momento de proceder con las diferentes máquinas a usar.

Además de tener en cuenta cómo los especialistas diagnostican el grado Gleason del tejido, es importante tener en cuenta el rol de la tinción del tejido pues esto puede ser útil o no para disponer de esa técnica de tinción y así extraer información relevante.

La comprensión de cada etapa y su rol en las máquinas superficiales es relevante pues puede representar un plan a seguir y una oportunidad para plantear de qué forma se quiere resolver el problema de clasificación, así como también el funcionamiento de los diferentes clasificadores tanto superficiales como profundos con la intención de sacar el máximo provecho para lograr un clasificador con buen desempeño en la tarea de clasificación con buena generalización y poca varianza al momento de diagnosticar imágenes microscópicas de tejidos de biopsias.

3. Descripción del dataset

El dataset fue obtenido a través del sitio web Kaggle, del desafío: “Prostate cANcer graDe Assessment” (PANDA) [8]. De esta manera, el desafío PANDA cuenta con un conjunto exactamente de 10616 imágenes de diapositivas completas de biopsias teñidas con H&E, las cuales fueron brindadas por el centro médico de la Universidad de Radboud junto con el Instituto de Karolinska, como se puede ver en la Tabla 3-1.

Tabla 3-1. Contenido del dataset escogido.

Centro Médico	Puntuación de Gleason	Cantidad de Imágenes
Karolinska	0+0	1925
	3+3	1814
	3+4	667
	3+5	13
	4+3	318
	4+4	466
	4+5	208
	5+3	2
	5+4	27
	5+5	16
Total Karolinska		5456

Radboud	3+3	852
	3+4	675
	3+5	67
	4+3	925
	4+4	660
	4+5	641
	5+3	41
	5+4	221
	5+5	111
	negative	967
Total Radboud	5160	
Total general	10616	

A su vez, es necesario destacar que todas las diapositivas de biopsias están escaneadas con un aumento de 20x, siendo archivo de imagen de tipo TIFF con formato RGB, teniendo como nombre un código de identificación ya que a través de la base de datos del archivo plano brindado train.csv, se encuentra el nombre de la imagen, junto con la clasificación de la puntuación de Gleason y gravedad de todo el tejido, para todos los centros médicos.

Por otra parte, se destaca que casi todas las imágenes del conjunto de entrenamiento tienen una máscara asociada, las cuales indican directamente qué partes del tejido están sanas y cuáles son cancerosas. Esta información de las máscaras es diferente en los 2 centros médicos.

Para el centro médico de Radboud, se identifica principalmente las glándulas prostáticas, con los siguientes valores:

Tabla 3-2. Contenido set de imágenes de la universidad de Radboud.

Etiqueta	Significado
0	Fondo (no tejido) o desconocido
1	Estroma (tejido conjuntivo, tejido no epitelial)
2	Epitelio sano (benigno)
3	Epitelio canceroso Gleason 3
4	Epitelio canceroso Gleason 4
5	Epitelio canceroso Gleason 5

Teniendo en cuenta estos valores de etiquetas de la Tabla 3-2, permite visualizar las zonas afectadas, señalando así la puntuación del patrón de crecimiento del cáncer en una zona específica del tejido, tal y como se observa en la Ilustración 3-1.



A. Tejido Benigno

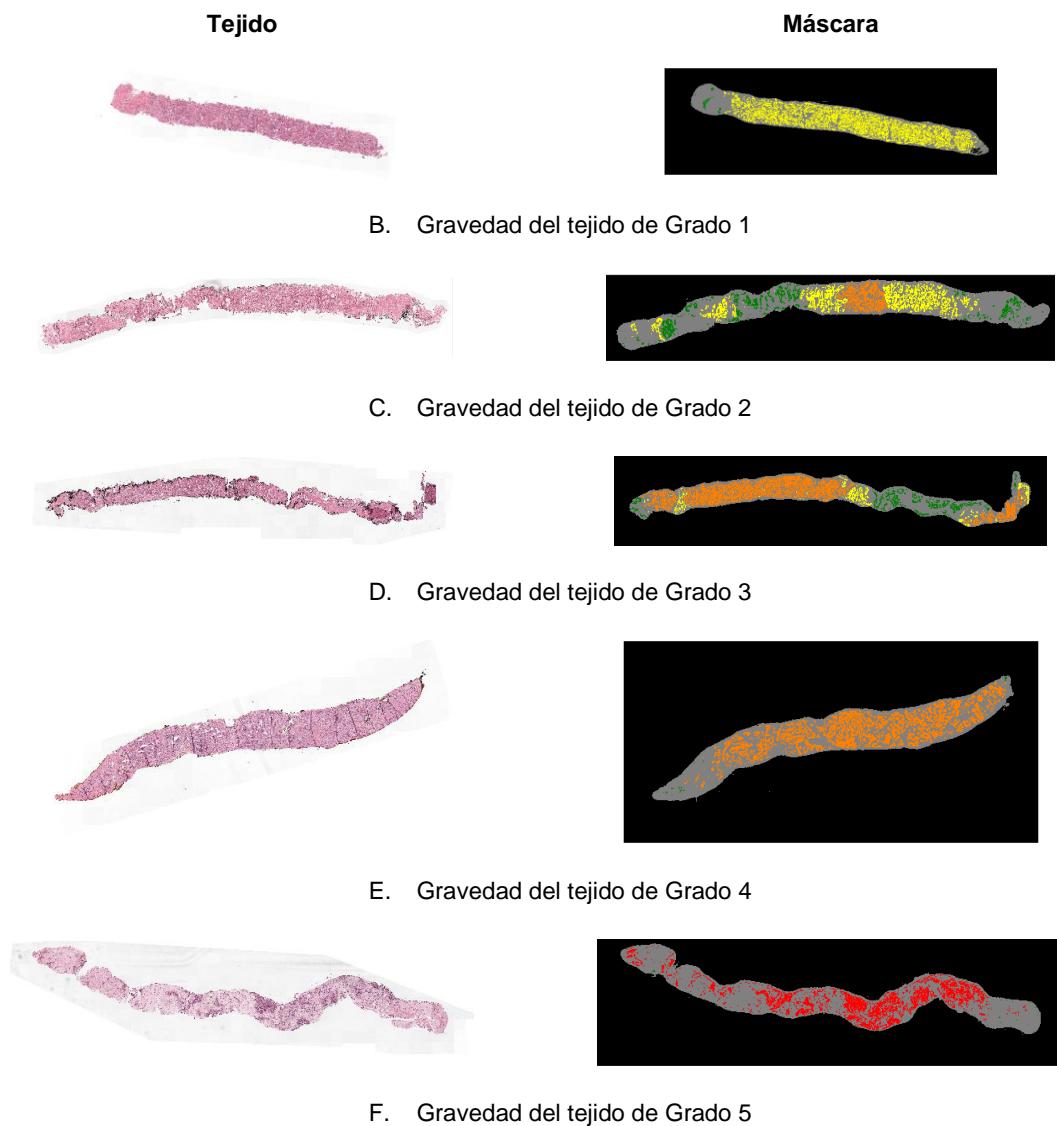


Ilustración 3-1. Tejidos con su respectiva máscara según la gravedad completa del tejido

Con la Ilustración 3-1, se puede visualizar de manera detallada las distintas zonas con presencia de cáncer, de acuerdo con el sistema de puntuación de Gleason, en donde la puntuación Gleason 0 señala una zona del tejido benigna, la cual se puede visualizar dentro de estas imágenes con color verde. Para el caso de una zona marcada con Gleason 3, es señalada con color amarillo, del mismo modo para zonas marcadas con color naranja representa que esa parte del tejido tiene una puntuación Gleason 4, y finalmente para zonas de Gleason 5, fueron representadas con color rojo.

Por otro lado, para el caso de Karolinska, se identifica sólo si el tejido es canceroso o benigno, teniendo las siguientes etiquetas:

Tabla 3-3. Contenido set de imágenes del instituto de Karolinska.

Etiqueta	Significado
0	Fondo (no tejido) o desconocido
1	Tejido benigno (estroma y epitelio combinados)
2	Tejido canceroso (estroma y epitelio combinados)

Estos valores de etiquetas de la Tabla 3-3, nos permiten visualizar que partes del tejido se encuentra sano, y cuales tienen cáncer. Un problema de este conjunto de imágenes es que no señalan específicamente el grado de Gleason, esto se evidencia en la Ilustración 3-2.



Ilustración 3-2. Gravedad del Tejido de Grado 2, con su tejido y máscara

En la Ilustración 3-2, se visualizan dos zonas, una zona de color verde que representa que partes del tejido son benignas, y la zona roja marca aquellas zonas con cáncer.

De este modo, para el desarrollo de este trabajo de grado, es necesario hacer uso de las imágenes del centro médico de Radboud, debido a que las etiquetas clasificadas cuentan con la puntuación Gleason, las cuales son necesarias para hacer las respectivas clasificaciones y categorizaciones de las diferentes características extraídas por cada parche.

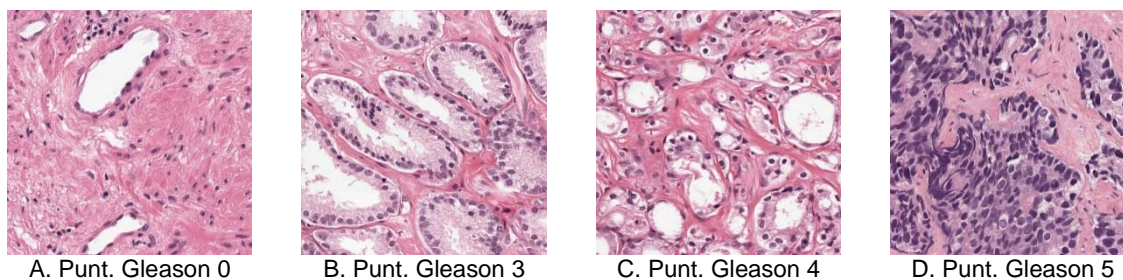


Ilustración 3-3. Parches con diferentes puntuaciones Gleason

En la Ilustración 3-3, se observan las diferentes puntuaciones de Gleason a distinguir. Esto es tarea de las siguientes secciones, en las cuales se prueban técnicas de preprocesamiento con el fin de resaltar la estructura y componentes del tejido, tales como: núcleos, glándulas y lumen del tejido. Estos elementos son partes importantes del tejido a distinguir en el campo de la Micrografía, como se explica en la sección 2.2.1.5.

Finalmente, para el entrenamiento y test de los diferentes clasificadores, se tomaron 409 tejidos prostáticos de los 5160 imágenes brindados por el centro médico de la universidad de Radboud. En donde, se extrajeron parches de dimensiones de 256x256 y 512x512, con las siguientes cantidades:

- Para los parches de 256x256, se extrajeron 5000 por cada clase (Punt. Gleason 0, 3, 4 y 5), siendo un total de 20000 parches. En donde se utilizaron 16000 para entrenamiento (4000 por clase) y 4000 para test (1000 por clase).
- Para los parches de 512x512, se extrajeron 1000 por cada clase (Punt. Gleason 0, 3, 4 y 5), siendo un total de 4000 parches. En donde se utilizaron 3200 para entrenamiento (800 por clase) y 800 para test (200 por clase).

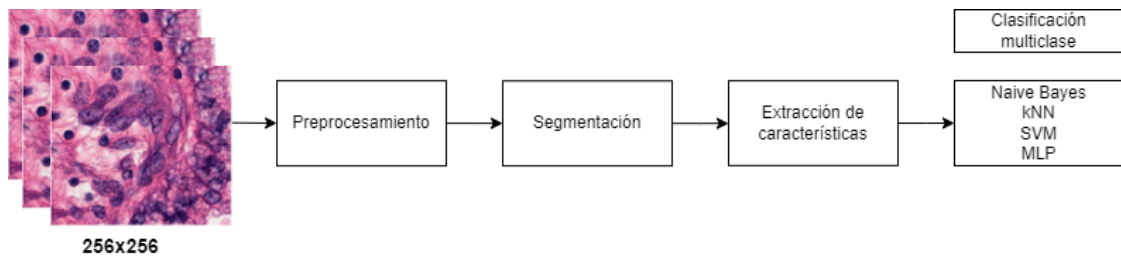
4. Metodología de la solución propuesta: Software

4.1. Introducción

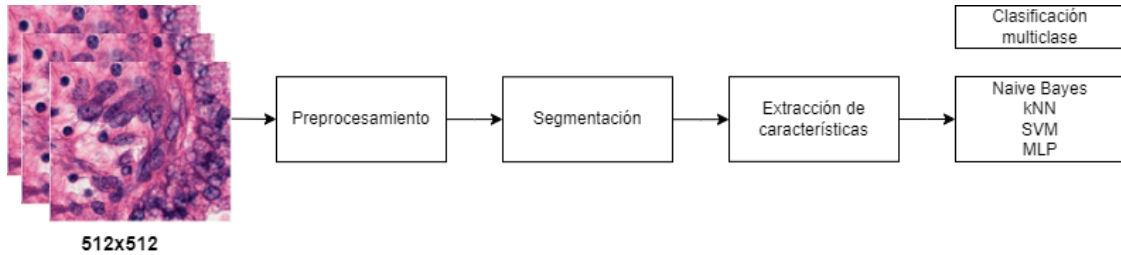
A continuación, se describe cada una de los procedimientos y técnicas utilizadas de las etapas del sistema con máquinas de aprendizaje superficial y del sistema con máquinas de aprendizaje profundo. El sistema propuesto tiene dos niveles de clasificación. El primero realiza la clasificación de pequeñas regiones de interés (parches) dentro de una imagen. El segundo estima la gravedad del cáncer en la imagen total a partir de la clasificación de los parches individuales.

El procedimiento con las máquinas de aprendizaje superficial comprende diferentes etapas las cuales pueden ser: Preprocesamiento, Segmentación, Extracción de características y la etapa final de clasificación.

Se hace hincapié en que las etapas pueden ser las mencionadas anteriormente pues se llevó a cabo diferentes arquitecturas en donde hay más o menos etapas de las mencionadas con el fin de probar que conjunto de etapas funciona de la mejor manera en la tarea de clasificación para hacer el diagnóstico del cáncer de próstata. En adición a la variación de arquitecturas, también se hizo variación en el tamaño de los parches a evaluar, tal y como se observa en la Ilustración 4-1, así como también la consecución de las diferentes etapas que se pueden implementar.



a. Diagrama de bloques para imágenes de tamaño de 256x256 píxeles



B. Diagrama de bloques para imágenes de tamaño de 512x512 píxeles.

Ilustración 4-1. Diagrama de bloques metodología de solución con máquinas superficiales- clasificación por parches.

Con las máquinas de aprendizaje profundo el procedimiento de clasificación de los parches se lleva a cabo mediante el uso también de los dos tipos de tamaño de los parches, a diferencia que para estas máquinas no es necesario las etapas de preprocesamiento, segmentación y extracción de características.

La segunda solución propuesta consiste en la clasificación de los tejidos por completo, lo que significa en un sistema no solamente capaz de clasificar un conjunto de parches entre los puntuaciones de Gleason 0,3,4 y 5, sino capaz de tomar una imagen de un tejido completo y dar como resultado el grado de severidad definitivo de todo el tejido en la escala de 0 a 5, como se puede evidenciar en la Ilustración 4-2.

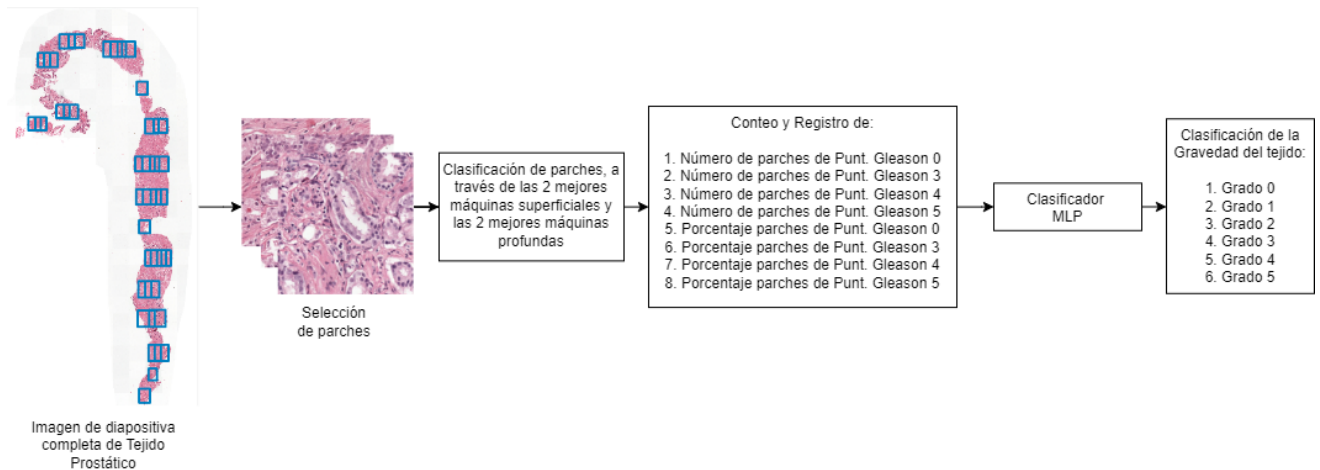


Ilustración 4-2. Diagrama de bloques metodológica de solución – clasificación tejido completo.

De esta forma se establecen dos formas de llevar a cabo la clasificación del grado de severidad de las imágenes histopatológicas de cáncer de próstata. La primera tal como se mencionó consiste en las evaluaciones de las imágenes en los tamaños 256x256 y 512x512 píxeles y una vez hecho esto se hará la evaluación de tejidos por completo con los clasificadores que demuestren los desempeños más altos al momento de realizar las pruebas.

4.2. Preprocesamiento

Para el preprocesamiento se tienen técnicas como mejoramiento de imágenes, ya sea ajustar sus niveles de grises, su composición de color de los 3 canales R, G y B, así como también el image patching por medio de recortes a la imagen original a un tamaño a elección propia. De esta manera, a continuación, se explican las diferentes etapas para la obtención de los parches y el mejoramiento de estos a través de diferentes técnicas, en donde se establecen sus parámetros de configuración.

4.2.1. Obtención de los parches

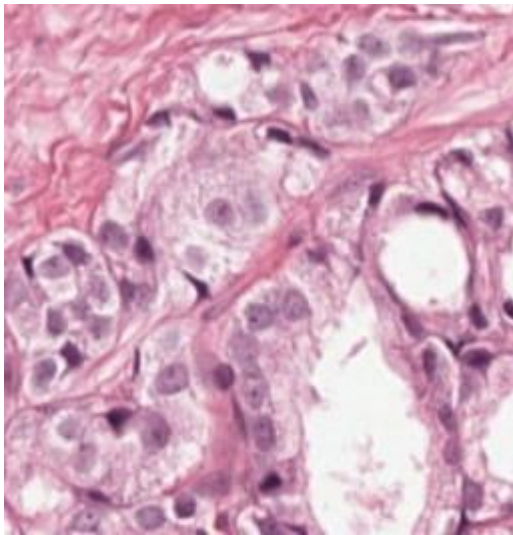
El dataset brindado por Kaggle [8], contiene imágenes de tipo WSI (Whole Slide Images), caracterizadas por su gran tamaño, es decir, representan un consumo de espacio considerable de memoria, tanto para el almacenamiento como para el tratamiento de estas por software. Este aspecto representa un desafío debido al considerable consumo de recursos computacionales necesario para su manejo, especialmente al trabajar con múltiples de estas imágenes simultáneamente. Además, la naturaleza detallada de las WSI conlleva una limitada extracción de información relevante, debido a que una característica de estas imágenes es que, si bien pueden tomar a un buen nivel de detalle del tejido, contienen demasiado fondo blanco, lo que dificulta la distinción del contenido crucial, pues es información que no aporta a la tarea de clasificación, demanda carga computacional y hace necesario dedicar una etapa para organizar estos datos en imágenes que aporten información sobre el problema y no fondo blanco con información vacía.



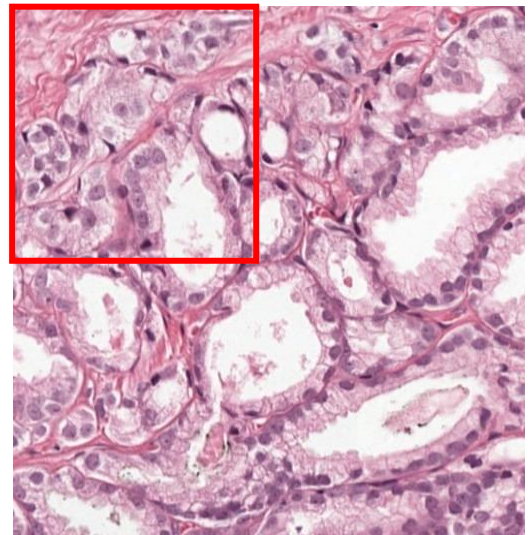
Ilustración 4-3. Imagen WSI de ejemplo de la base de datos que se dispone.

Con el fin de trabajar con la información relevante y que aporta al tratamiento de los datos para que posteriormente sea procesado y obtenido para la clasificación se opta por realizar

un reajuste en la dimensión de las imágenes, logrando así un aumento de datos e información a tratar y clasificar. Esto se hace porque como se ve en las Ilustración 4-3, la proporción de fondo blanco respecto a la imagen es mayor y no aporta para la tarea de clasificación.



A. 256x256 pixeles



B. 512x512 pixeles

Ilustración 4-4. Image patching de las imágenes del dataset para una mejor representación de la información.

Como se observa en la Ilustración 4-4 al llevar a cabo image patching se logra concentrar la información de mejor forma por imagen, siendo “A” una imagen a un tamaño de 256x256 píxeles y “B” de 512x512 píxeles, en donde el recuadro rojo representa la imagen “A” dentro de “B”, evidenciando así que “B” al ser de mayor tamaño logra una mayor expansión de la información que se obtiene con las imágenes de la imagen en “A”. Si bien este aumento de tamaño puede resultar en una disminución en la cantidad de parches sacados de un tejido prostático, garantiza que se puede observar de mejor manera la forma de muchas glándulas dentro de la imagen, lo cual es una opción de posible mejora a considerar y por ello se dividen los datos en estos dos conjuntos de tamaño para evidenciar así, cuáles son las pruebas con mejores desempeños.

También cabe resaltar que debido a la tinción que se le hace este tipo de tejidos para resaltar los núcleos de las células y demás partes del tejido, que los especialistas le brindan importancia y que resulta en zonas del tejido con coloración no deseada como coloración de color verde, roja y azul, tal y como se puede apreciar en la Ilustración 4-5.

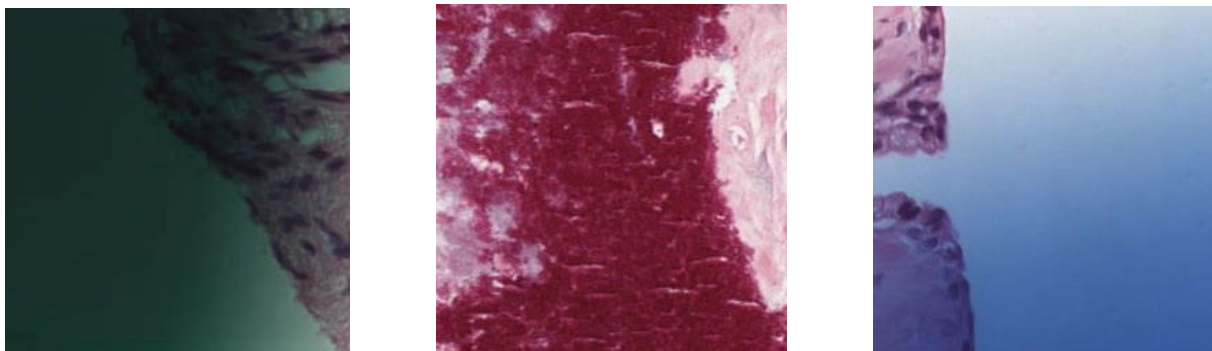


Ilustración 4-5. Presencia de tinción no deseada en los tejidos.

Teniendo en cuenta esto, además de realizar image patching es necesario realizar un proceso de selección de los parches, basado en la media y desviación estándar de los píxeles de los colores no deseados presentes en los tejidos para reducir este tipo de imágenes con ruido, pues si se permiten esa serie de imágenes puede alterar el proceso obtención de características y a su vez traer problemas en la clasificación. Siendo así, este proceso de selección se hace mediante los valores de los intervalos en la Tabla 4-1.

Tabla 4-1. Intervalo de valores para selección de parches sin ruido por tinción no deseada.

Características	Relación: tejido/fondo		Canal R		Canal G		Canal B	
	Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo
Media	150	198	167	239	115	185	144	200
Desviación estándar	22	57	12	54	23	64	16	49

Los rangos de la Tabla 4-1 se establecieron mediante la observación del comportamiento de las características en las imágenes sin ruido, es decir, se escogieron imágenes como la Ilustración 4-4, que no presentaban sobrecoloraciones, en las que se obtuvieron la media y desviación estándar, estableciendo así un intervalo de aceptación para cada canal R,G,B de las imágenes. Finalmente, este proceso se puso a prueba con imágenes que si presentaban sobrecoloraciones, como las que se encuentran en la Ilustración 4-5, en donde se encontró que estas imágenes fueron efectivamente rechazadas, debido a que sus características de media y desviación estándar no se encontraban dentro del rango establecido para los 3 canales.

Se usan los valores como media y desviación estándar pues estos describen la tendencia de color dominante en una imagen (media) y qué tan alta es la variación de los colores presentes en la imagen (desviación estándar), de esta forma se evitan parches con las variaciones del color debido a la tinción presente en el tejido y así evitar el ruido de la información agregada que no aporta a nada a la representación de los tejidos.

Una vez se lleva a cabo este proceso, se aplican diferentes técnicas de preprocesamiento para ajustar las imágenes aún más y tomar diferentes regiones de interés o hacer que las

regiones de interés se realcen por diferentes formas, es decir, un realce de la región de interés a través de diferentes conversiones de escala de grises, a través de la aplicación de una segmentación a la imagen y demás técnicas existen en visión artificial.

4.2.2. Conversión a escala de grises

Una técnica aplicada a las imágenes con las que se trabajó fue la conversión a escala de grises, pues toda imagen a color es conformada por 3 canales principales por el modelo que en general se usa que es el RGB. Por facilidad en el tratamiento de los datos, siempre que se trabaja con imágenes cada píxel es una característica y el tener 3 canales hace que los datos a tratar aumenten, en consecuencia, a esto una técnica aplicada es pasar la dimensionalidad de los datos de 3 a 1 dimensión y hacer más fácil el tratamiento de los datos.

4.2.2.1. Corrección gamma

Se aplica esta técnica con el fin de ajustar el contraste entre los núcleos presentes en el tejido con respecto al resto de este ajustando el valor de Gamma, tal como se observa en la Ilustración 4-6.

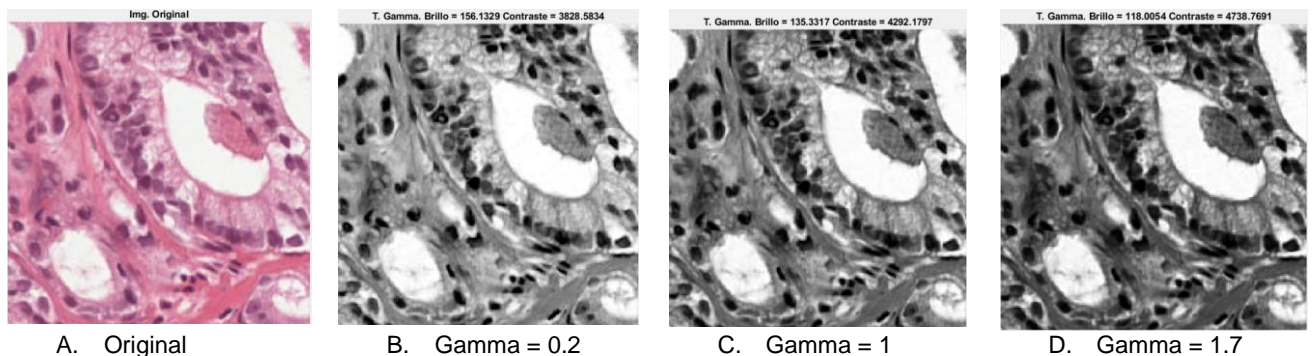
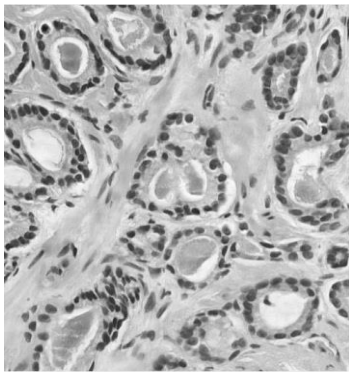


Ilustración 4-6. Efecto de aplicar conversión escala de grises y corrección Gamma.

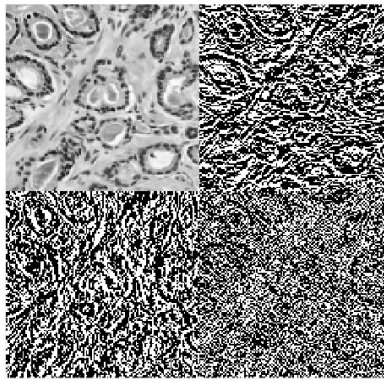
A partir de la variación del Gamma, se determinó que con un Gamma = 1, había un buen contraste y brillo, en donde se resaltaban perfectamente los núcleos y las glándulas del tejido prostático, estableciendo este parámetro para todos los parches extraídos.

4.2.2.2. Transformada Wavelet

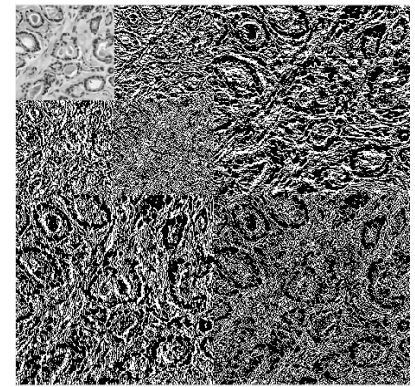
Esta técnica también se aplica para resaltar los núcleos y establecer el contorno o regiones que conforman la glándula del tejido prostático, tal y como se observa en la Ilustración 4-7. Cabe destacar, que se usó una descomposición en 2 niveles de wavelet, usando los filtros Haar, en donde se tiene un filtro pasa bajas $Lo = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]$ y un filtro pasa altas $Hi = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right]$.



A. Original



B. Transformada de Wavelet de nivel 1

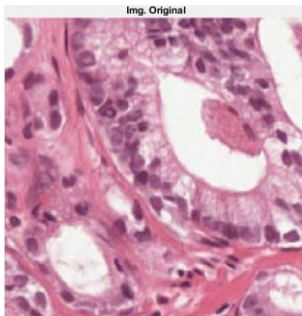


C. Transformada de Wavelet de nivel 2

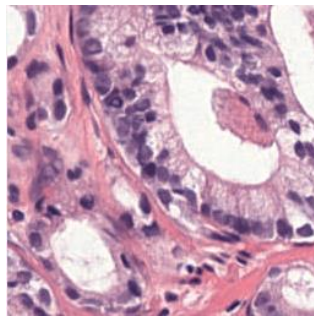
Ilustración 4-7. Aplicación de la transformada Wavelet multi nivel a las imágenes.

4.2.3. Preprocesamiento a color

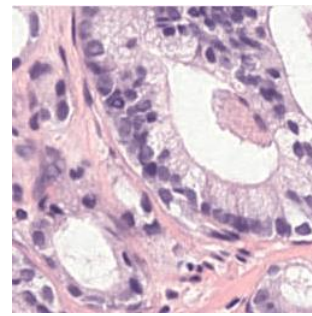
El preprocesamiento a color que se hizo en esta etapa fue la Normalización de la tinción, (Staining Normalization) para imágenes de tinción con colorantes químicos H&E. Esto se hace teniendo en cuenta la coloración que realizan los especialistas a los tejidos con los químicos de hematoxilina y eosina (H&E) con el fin de que en estos tejidos se resalten características para analizar mejor estos tejidos y hacer el diagnostico más fácil, en donde la tinción por medio de H (hematoxilina) permite resaltar las células del tejido prostático, y la eosina (E) el tejido celular. De esta manera, el objetivo de la normalización de la tinción es regular la intensidad y concentración de los químicos de hematoxilina (H) y eosina (E), permitiendo resaltar la estructura del tejido o componentes celulares que se encuentran dentro de este mismo. [50]



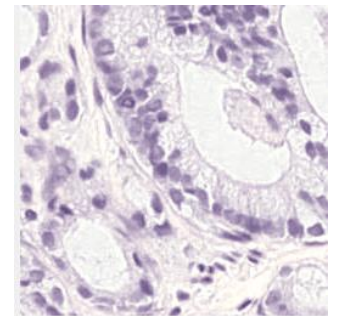
A. Original



B. $maxCRef = [1.97; 1.03]$



C. $maxCRef = [1.5; 0.5]$



D. $maxCRef = [1.1; 0.1]$

Ilustración 4-8. Manipulación del color en los canales H y E.

En el proceso de preprocesamiento de imágenes a color, se emplea la normalización de la tinción. Este procedimiento implica principalmente reducir las variaciones causadas por la tinción con eosina (E), enfocándose así en los efectos producidos por la tinción con hematoxilina (H) [50]. Este enfoque se logra ajustando el parámetro $maxCRef$, detalle que se puede apreciar claramente en la Ilustración 4-8. La razón detrás de esta técnica es que las características destacadas por la tinción con H son cruciales para identificar patrones de crecimiento asociados a la agresividad del cáncer. Los núcleos, que se resaltan más con la

tinción H, son clave para discernir entre los distintos grados de severidad del cáncer de próstata. Por lo tanto, se minimiza la influencia del canal E, que, aunque realza mejor el tejido celular, no contribuye significativamente a diferenciar los niveles de severidad de esta enfermedad, usando así imágenes como la Ilustración 4-8-D.

También se considera dentro del procesamiento, el cambio del modelo de color de RGB a HSV [51], pues en este modelo se presenta la facilidad de disminución ante cambios de iluminación, variabilidad menor entre la tinción del tejido y facilita el proceso de segmentación posterior en caso de que se desee hacer una etapa de segmentación, pues en función del color y saturación canales V y S respectivamente, permite resaltar fácilmente la distinción entre diferentes regiones.

4.3. Segmentación

La segmentación (ver apartado 2.2.3) tiene como función continuar el realce de la región de interés que se viene haciendo desde la etapa de preprocesamiento, para este trabajo se utilizó la Segmentación basada en umbrales y regiones, y otra segmentación basada en color por clustering.

4.3.1. Segmentación basado en umbrales y regiones

Para la segmentación basada en regiones, se utilizó el método Watershed, el cual fue escogido a través de una comparación de rendimiento entre el método Otsu y contornos activos por medio de la obtención de los índices de Jaccard, Dice y BF score y contrastando ese rendimiento con ground truth hechos a través de la herramienta “Image Segmenter” de Matlab.

Tabla 4-2. Índices de calidad de la segmentación de los 3 métodos de segmentación.

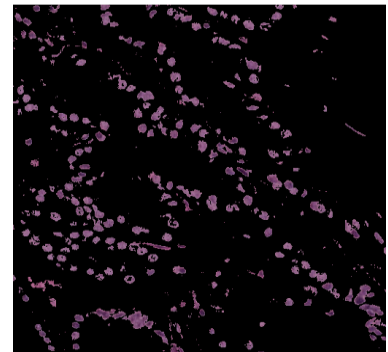
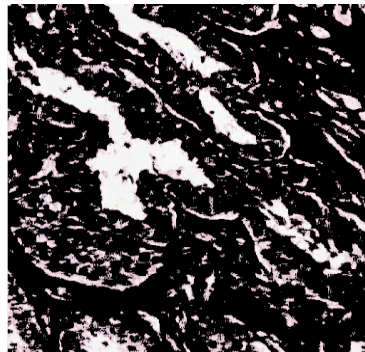
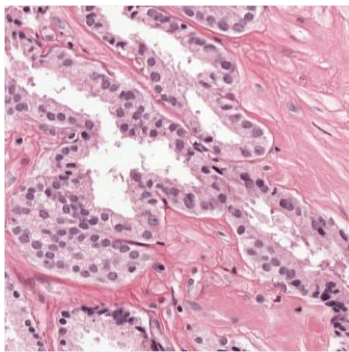
	Otsu vs Ground Thruth	Contornos activos vs Ground Thruth	Watershed vs Ground Thruth
Jaccard Index	0.8181	0.8094	0.8209
Dice Index	0.8994	0.8671	0.9010
BF score	0.9353	0.9535	0.9358

Tal como se observa en la Tabla 4-2, se distingue que el método Watershed es buen método según se observa en los índices de calidad de la segmenación, para realizar las segmentaciones de los núcleos donde se obtiene un valor 0.82 para el índice de Jaccard, 0.90 para el índice de Dice y 0.9358 para el BF score. Siendo los índices de Jaccard y Dice siempre los más altos comparado con el método Otsu y Contornos activos.

4.3.2. Segmentación basada en colores por clustering

Para realizar la segmentación basada en colores por clustering, a través del agrupamiento K-means, es necesario establecer 3 clústeres ($k = 3$), segmentando así el estroma, el lumen y los núcleos presentes en el tejido. En donde, para utilidad del trabajo se utilizó solo el lumen y los núcleos en las diferentes imágenes H&E, tal y como se observa en la Ilustración 4-9. El

estroma no se toma en cuenta como se mencionó anteriormente en la sección 4.2.3, debido a que es una característica que no forma parte del modelo de clasificación Gleason para el diagnóstico.



A. Original

B. Lumen

C. Núcleos

Ilustración 4-9. Aplicación del algoritmo k means en las imágenes.

Como se puede observar en la Ilustración 4-9, la aplicación que se hace es la de k-means, en donde se observan las dos segmentaciones, la imagen B representa la región de interés que es el lumen y la imagen C representa la segmentación de los núcleos presentes en el tejido.

4.4. Extracción de características

4.4.1. Por contorno

En la extracción de características por contorno se extrae información con base en la forma de las células, de tipo morfológicas [52]. Estas características lo que describen atributos como el centroide, longitud de los ejes mayor y menor de los núcleos de las células, como tal representan la forma en cómo puede estar presente los núcleos o glándulas en una célula. Se aplicó a las imágenes segmentadas y en total se extrajeron 9 características:

Tabla 4-3. Características de tipo morfológicas extraídas.

Tipo de características	Características extraídas
Morfológicas	Área, perímetro, área cóncava, diámetro equivalente, excentricidad, solidez, longitud de extensión del eje mayor, longitud del eje menor.

4.4.2. Por textura y momentos estadísticos

La extracción de características por textura se basa en obtener información representativa de los tejidos de forma estadística. El objetivo es sacar medidas estadísticas que establezcan las relaciones o asociaciones, así como discrepancias entre los pixeles que describen cada clase, ya sea Gleason 0, 3, 4 o 5.

Para las arquitecturas con las que se decidió trabajar con características de textura de las imágenes se les extrajo la matriz de coocurrencia de grises o GLCM (Gray Level Co-occurrence Matrix) [37], estableciendo que la distancia a examinar entre pixel y su vecino de

interés es unitaria ($d = 1$) evaluando esto en las direcciones 0° , 45° , 90° y 135° . De esta manera, a partir de la GLCM obtenida se calcularon 21 características de textura para el conjunto de imágenes con el que se trabajó, las cuales son:

Tabla 4-4. Características de tipo textura extraídas a las imágenes.

Tipo de características	Características extraídas
Textura	Autocorrelación, contraste, correlación, prominencia de grupo, sombra de grupo, disimilitud, energía, entropía, diferencia inversa, homogeneidad, probabilidad máxima, suma de cuadrados, suma de la media, suma de la varianza, suma de la entropía, diferencia de la varianza, diferencia de la entropía, medida de la correlación 1, medida de la correlación 2, diferencia inversa normalizada y diferencia de momento normalizada.

Esta matriz se decidió escoger como vector de características pues en los antecedentes presentó un uso frecuente y como características de tipo textura también se encontró en el rango de resultados de clasificación altos.

Además, también se extrajo características de momentos de color en cada canal R, G y B. Cinco por cada canal:

Tabla 4-5. Características estadísticas extraídas por canal R, G y B.

Tipo de características	Características extraídas por canal R, G y B.
Estadísticas	Media, varianza, desviación estándar, oblicuidad y curtosis

4.4.3. Selección de características

Se hizo una selección de características a través de dos técnicas:

La primera técnica fue a través de un ranking por medio del algoritmo reliefF que arroja como resultado un orden de características de mayor a menor importancia desde el primer índice hasta el último respectivamente. Una vez con este ranking lo que se hizo fue variar manualmente la reducción de las características pues reliefF da un ranking, pero no elimina o proporciona un grupo de características idóneo. [53]

La segunda técnica fue a través de un método embebido en donde se prueban aleatoriamente conjuntos de características y se hace la clasificación con esos conjuntos, cada conjunto se prueba con un clasificador dependiendo del método, para este caso el clasificador usado fue Random Forest, pues la capacidad de generar arboles de decisión permite probar diferentes combinaciones y evaluar qué características funcionan mejor para la clasificación y el conjunto de características a usar se eligió a través de una variación de diferentes conjuntos de características escogiéndolo el grupo que mejor desempeño permite alcanzar. [53]

4.5. Clasificación

Una vez realizado todo el trabajo en las etapas anteriores se procede con la clasificación. Cabe resaltar, que lo realizado anteriormente es para las máquinas superficiales pues son a estas en donde ocurre la aplicación de dichas técnicas de visión artificial.

La clasificación que se realiza es como se describe en la Ilustración 4-1, se utilizan imágenes de 256x256 y 512x512 pixeles.

4.5.1. Clasificadores - Máquinas superficiales

Para realizar la clasificación se llevan a cabo diferentes pruebas, para un total de 5 pruebas:

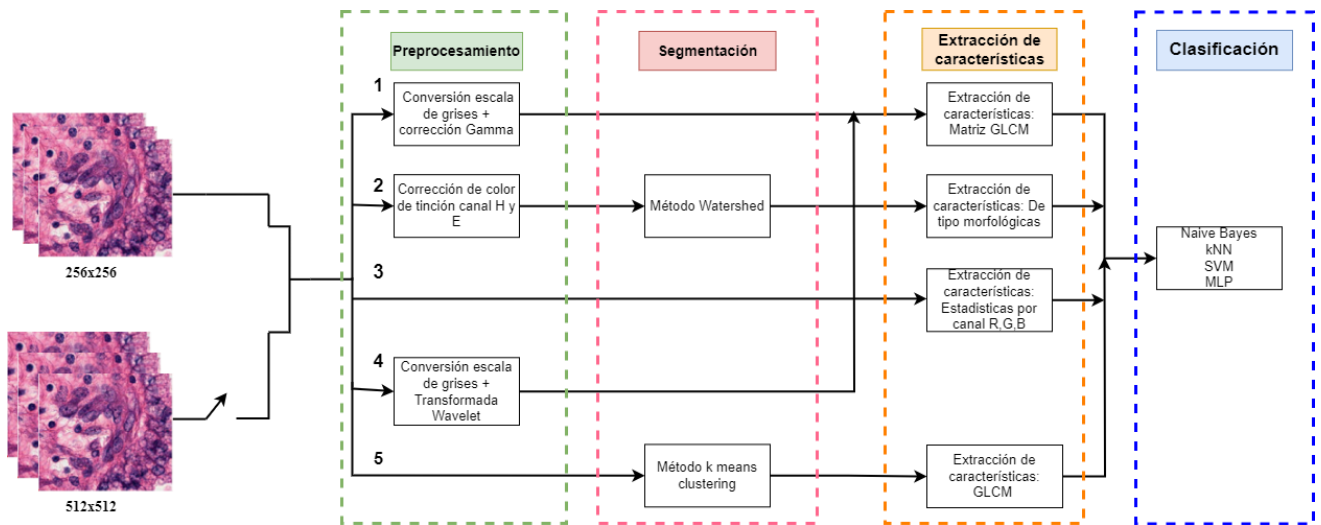


Ilustración 4-10. Pruebas realizadas con las máquinas superficiales.

En la Ilustración 4-10, en primer lugar se destaca que para las imágenes de 256x256 píxeles se realizan las siguientes 5 pruebas diferentes:

1. **Prueba 1:** Para esta prueba las imágenes de tinción H&E, que están en formato RGB se convierten a escala de grises para luego aplicar una corrección gamma, con los detalles indicados en la sección 4.2.2.1. Esta arquitectura se estableció mediante lo visto en las electivas profesionales de visión artificial y reconocimiento de patrones, considerando solo los bloques de preprocesamiento, extracción de características por textura y clasificación.
2. **Prueba 2:** Esta prueba consiste en hacer uso de la normalización de la tinción, resaltando los núcleos a partir de dejar el aporte de la tinción H, como se detalla en la sección 4.2.3. Esta arquitectura se apoya en utilizar el método de segmentación de Watershed, para extraer las características morfológicas de los núcleos, siendo esta prueba basada en lo hecho por el antecedente [14] consultado, respecto a la forma en cómo usan el método de Watershed y se extraen esas mismas características que se exponen dentro de este mismo artículo.

3. **Prueba 3:** En este intento, no se aplica ninguna técnica de preprocesamiento, con la intención de usar las características que ofrece la imagen a color, sin ser manipulada. Para esta arquitectura, se extraen características de tipo estadísticas, por cada canal del modelo de color de la imagen R, G y B. Esta prueba se llevó a cabo de esta manera influenciada por lo hecho por los investigadores en el antecedente [13].
4. **Prueba 4:** Para este nuevo intento, se aplica una nueva técnica de preprocesamiento como es la transformada Wavelet multinivel, recomendada en el artículo [13] y diferente a la técnica en la prueba 3, siendo útil para resaltar las glándulas presentes en el tejido. En adición al bloque de preprocesamiento, se extraen características de textura por medio de la matriz GLCM, utilizado en los artículos [12],[13] y [15].
5. **Prueba 5:** Para este ensayo, se utilizaron las técnicas de segmentación basada en colores por clustering a las imágenes en donde se extraen el lumen y núcleo en dos imágenes por separado. A estos dos grupos de imágenes se les hace una conversión a escala de grises y una conversión del modelo de color de RGB a HSV para posteriormente descomponer las imágenes en formato HSV en 3 canales, H, S y V, para finalmente extraer las características de estas imágenes por medio de la matriz GLCM, proceso que se puede observar en la Ilustración 4-11. Esta arquitectura se basa en el trabajo hecho por los autores en el artículo [15].

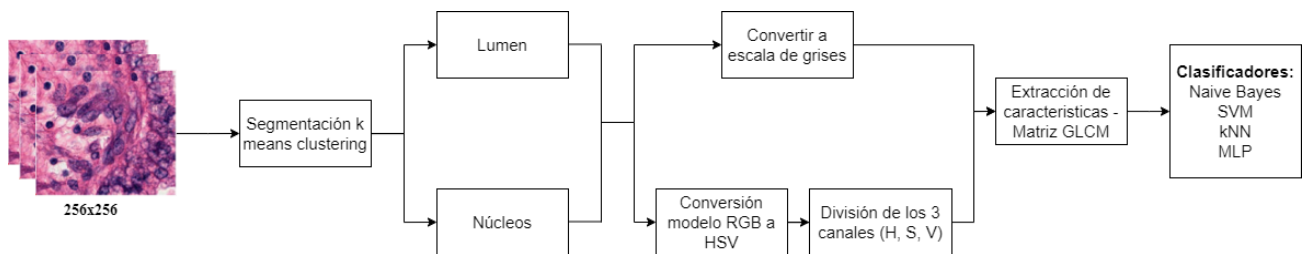


Ilustración 4-11. Diagrama de bloques de la prueba 5.

Una vez terminado el proceso de evaluación de las 5 pruebas con el conjunto de imágenes de 256x256 píxeles, se revisan cuáles son las 2 mejores técnicas y los 2 mejores clasificadores en función de sus desempeños, para así evaluar el conjunto de imágenes de 512x512 píxeles con este conjunto de técnicas y clasificadores seleccionados. Esto se realiza esperando aumentar los resultados obtenidos con las imágenes de tamaño 256x256.

4.5.2. Clasificadores - Máquinas profundas

Las pruebas realizadas con las máquinas profundas involucran 3 arquitecturas de redes neuronales convolucionales: AlexNet, ResNet50 y EfficientNetB0. Se escogieron estas redes pues se encontró que son redes útiles en este campo de aplicación [11], [16].

Debido a que las redes neuronales o máquinas profundas tiene la capacidad de extracción y selección de características relevantes por ellas mismas y no es necesario que el investigador haga esas tareas, termina siendo un requerimiento principal un conjunto de

datos grandes y balanceados para evitar el sobre ajuste u overfitting como se conoce en inglés.

Para evitar el sobreajuste en los datos, se optó por emplear modelos pre-entrenados, es decir redes neuronales convolucionales con pesos establecidos en tareas de reconocimiento de imágenes. Esta técnica es conocida como "transfer learning", implica adaptar un modelo desarrollado para una tarea específica, para que funcione en una tarea diferente pero relacionada, en este caso, la clasificación de imágenes de biopsias de próstata. Adicionalmente, se experimentó con imágenes de dos resoluciones diferentes, 256x256 y 512x512 píxeles, para evaluar cómo la variación en la resolución de las imágenes afecta el rendimiento de estos modelos en la clasificación de la severidad del cáncer de próstata.

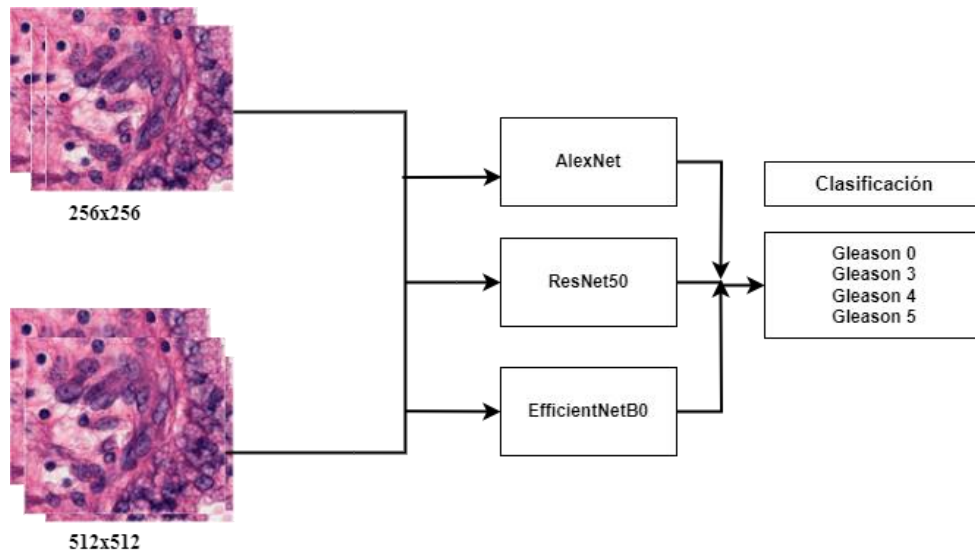


Ilustración 4-12. Clasificación a través de las máquinas profundas.

4.6. Evaluación del tejido completo

Una vez se han hecho las pruebas para todas las máquinas superficiales y profundas la segunda solución propuesta consiste en tomar los dos mejores clasificadores superficiales y profundos y evaluar los tejidos completos (Ilustración 4-2), comparar sus resultados y ver cuál o cuáles de los cuatro responde mejor a la tarea de clasificación del tejido completo.

4.7. Conclusiones

Con estas soluciones propuestas lo que se busca siempre es que los clasificadores logren la capacidad de diferenciar bien entre las diferentes clases de la puntuación Gleason, aplicando técnicas que proporcionen una separación de la región de interés al resto de tejido y características discriminantes para diferenciar las clases de forma satisfactoria por lo que se realizan todas las técnicas registradas en esta sección de diferente forma y en diferentes conjuntos de grupos de técnicas agrupadas buscando esa diferenciación y generalización del problema de clasificación.

Se reconoce el aporte al tratamiento de datos en la etapa de preprocesamiento pues la obtención de parches potencia la capacidad de primero tener información relevante en vez de solamente un tejido completo que a grandes rasgos no se puede analizar a diferencia si se centra la atención en zonas donde se vea más de cerca el tejido, también se reconoce la capacidad de remoción del ruido existente en los parches debido al uso de diferentes químicos para llevar a cabo la tinción de estos tejidos, pues en un primer momento no basta con realizar la división de la imagen por parches pues si no se tiene en cuenta la presencia de este ruido puede resultar en problemas de obtención de características poco representativas de los diferentes patrones y en una mala clasificación.

La etapa de segmentación por método Watershed si bien cumple su objetivo de separar las zonas de interés que en este caso se decidió que sean los núcleos del resto del tejido, se reconoce igual que en ocasiones resulta en una segmentación que no captura en su totalidad la región de los núcleos lo que puede resultar en un sesgo de cómo queda representada la información ofrecida por cada tejido, que puede provocar errores de clasificación si los vectores de características extraídos de esta etapa resultan pobres en información discriminativa de un puntaje Gleason a otro.

En la extracción de características hechas para este trabajo se destacan características de contorno, textura y momentos estadísticos. Para las características por contorno, se destaca por su utilidad en describir en términos de geometría la forma de las células, permitiendo así tener información sobre las células, que es un aspecto que se tiene en cuenta por los especialistas al momento de diagnosticar la severidad del adenocarcinoma de próstata. Por el lado de las características de textura, la utilidad por la que se decidió su uso es debido a la capacidad de describir la estructura superficial o apariencia de las regiones en una imagen, que es de suma importancia en este tipo de imágenes, lo que aporta de forma significativa a este trabajo en donde las regiones de interés son los patrones de crecimiento de las células y el cambio estructural de las glándulas, a su vez también resulta útil obtener los momentos estadísticos para cada canal R, G y B, análisis con capacidad de poder generar características con una discriminación marcada permitiendo evidenciar los cambios en las divisiones celulares y cambios en la apariencia estructural de las glándulas, gracias al análisis separado por cada canal.

La selección de características se obtiene mediante el uso de los algoritmos reliefF y Random Forest, que permiten tener los aspectos de características más importantes, en donde se evita obtener un espacio de características con alta dimensionalidad que genere problemas en los clasificadores impidiendo su capacidad de discriminación entre las clases. A su vez, esto permite reducir la carga computacional ya que se descartan características irrelevantes que demanda tiempo de procesamiento durante la clasificación.

Teniendo en cuenta todo lo anterior, detrás de cada técnica escogida hay un objetivo de la clasificación de la severidad del cáncer, no solo enfocándose en los clasificadores, sino también teniendo en cuenta cada etapa. En el caso de la etapa de preprocesamiento, el enfoque es aplicar técnicas que propicien una limpieza u optimización de la imagen para

eliminar primero efectos no deseados como ruido o coloración excesiva que no aporta a la distinción de la severidad del cáncer. Esto también está conectado a una etapa de segmentación que busca potenciar esos resultados hechos desde la etapa anterior, permitiendo así resaltar las células y glándulas que se encuentran dentro del tejido, mejorando de este modo la capacidad de discernir entre un grado de severidad u otro. Estos pasos en conjunto permiten una extracción de características capaz de contribuir a establecer un criterio de discriminación entre clases que permita a los clasificadores lograr altos rendimientos de clasificación. Aunque desde el punto de vista, en el campo de visión artificial y el reconocimiento de patrones, es una tarea de clasificación de múltiples clases, como esta tarea está relacionada con la salud de un paciente, es importante asegurarse de que la tarea de clasificación tenga como objetivo diferenciar un grado de gravedad de otro.

5. Pruebas y resultados

5.1. Introducción

Esta sección permite evaluar y comparar las diferentes técnicas de visión artificial y reconocimiento de patrones con las máquinas de aprendizaje superficial y profundo implementadas. Estas pruebas permitieron determinar y seleccionar las técnicas y sus respectivos parámetros que brindan mayor desempeño de clasificación de la puntuación Gleason de cada parche y del estadio del cáncer de las imágenes histopatológicas. Con base en el desempeño final se puede estimar la confiabilidad del diagnóstico e identificar los alcances y limitaciones del sistema desarrollado.

De esta manera se registran los resultados obtenidos para las máquinas de aprendizaje superficial y profundo teniendo en cuenta que se hace una clasificación en dos niveles, el primero es una clasificación multiclase y segundo la implementación de un sistema que se encarga de tomar una imagen de un tejido completo y diagnosticar el grado Gleason del tejido completo a partir de la selección del mejor clasificador del conjunto de pruebas.

5.2. Máquinas de aprendizaje superficial

5.2.1. Descripción

Este apartado contiene la ejecución de los resultados obtenidos al probar las diferentes arquitecturas con las máquinas superficiales propuestas Naive Bayes, kNN, SVM y MLP (ver sección 4.5.1). Se detalla cómo se llevó a cabo las pruebas y cómo se llegó a las dos máquinas superficiales para la clasificación multiclase de la puntuación Gleason.

5.2.2. Procedimiento: Imágenes con tamaño de 256x256 píxeles

Lo realizado acorde a la solución propuesta 1 (ver sección 4.5.1) consta de 2 pruebas, primero se prueban diferentes arquitecturas de solución (Ilustración 4-10), primero con imágenes de dimensión de 256x256 píxeles para reconocer aquellas técnicas y clasificadores que aporten de mayor manera a la clasificación de la puntuación Gleason, con el fin de generar una unión de técnicas ya sean de preprocesamiento, segmentación y extracción de características de distinto tipo para así ver si aumentando la resolución a 512 x 512 se logran

mejores resultados que con una resolución de 256x256 píxeles y a su vez para reconocer los mejores clasificadores entre Naive Bayes, kNN, SVM y MLP.

5.2.3. Resultados y análisis

Tabla 5-1. Resultados máquinas superficiales imágenes de 256x256 píxeles

Prueba	Etapas	Características	Clasificadores	Exactitud
1	<ul style="list-style-type: none"> - Preprocesamiento - Extracción de características - Clasificación 	21 de textura, reducidas a 10	<ul style="list-style-type: none"> - Naive Bayes - kNN - SVM - MLP 	NB: 0.399 kNN: 0.493 SVM: 0.516 MLP: 0.529
2	<ul style="list-style-type: none"> - Preprocesamiento - Segmentación - Extracción de características - Clasificación 	9 características de tipo morfológicas, reducidas a 4	<ul style="list-style-type: none"> - Naive Bayes - kNN - SVM - MLP 	NB: 0.297 kNN: 0.277 SVM: 0.288 MLP: 0.294
3	<ul style="list-style-type: none"> - Extracción de características - Clasificación 	15 tipo estadísticas, reducidas a 7	<ul style="list-style-type: none"> - Naive Bayes - kNN - SVM - MLP 	NB: 0.462 kNN: 0.519 SVM: 0.548 MLP: 0.562
4	<ul style="list-style-type: none"> - Preprocesamiento - Extracción de características - Clasificación 	126 de textura, reducidas a 63	<ul style="list-style-type: none"> - Naive Bayes - kNN - SVM - MLP 	NB: 0.382 kNN: 0.410 SVM: 0.434 MLP: 0.411
5	<ul style="list-style-type: none"> - Segmentación - Extracción de características - Clasificación 	168 características de tipo estadísticas + textura reducidas a 84	<ul style="list-style-type: none"> - Naive Bayes - kNN - SVM - MLP 	NB: 0.458 kNN: 0.554 SVM: 0.564 MLP: 0.628

Al entrenar y probar las 5 arquitecturas propuestas se obtienen en algunos casos resultados con bajo desempeño, por debajo del 50% como las pruebas 2 y 4 y en las demás si bien los desempeños superan el 50% de exactitud su resultado es poco deseable pues no representan una fiabilidad para el diagnóstico del cáncer de próstata, como se puede observar en la Tabla 5-1.

Sin embargo, en las pruebas 1, 3 y 5 se nota que los clasificadores SVM y MLP son los clasificadores que logran mejores resultados en cuanto a exactitud respecto a los demás clasificadores en todas las pruebas y además se nota que las características que se usan en las pruebas 3 y 5 son pruebas en donde se obtuvo el mayor desempeño de exactitud, obteniendo en la SVM un desempeño de 0.548 de exactitud en la prueba 3 y 0.564 de exactitud en la prueba 5. Para el caso de MLP, se evidenció un desempeño de 0.562 de exactitud en la prueba 3 y 0.628 de exactitud en la prueba 5.

Con base a estos resultados se observa pues que los conjuntos de características estadísticas a partir de los canales R, G y B, así como las características de textura junto con una segmentación con clustering ofrecen mayor aporte a la clasificación junto con SVM y MLP y son estos dos clasificadores los cuales se escogen para hacer una prueba en donde se mejore el tamaño de la imagen y con un método de selección de características distinto en donde se involucre la validación cruzada para encontrar una mejor distribución de los datos para entrenar y evaluar. Cabe resaltar, que se hace una constante reducción de dimensión del espacio de características con el fin de reducir la complejidad del mismo y evitar obtener clasificadores con sesgo en el diagnóstico.

5.2.4. Procedimiento: Imágenes de tamaño 512x512 pixeles

Con base en los resultados con las imágenes de 256x256 pixeles lo que se decide hacer es escoger los clasificadores MLP y SVM y hacer una unión de las características estadísticas extraídas a partir de cada canal de color del modelo RGB y las características que se obtienen a través de aplicar k means clustering junto con la conversión de modelos de color, tal como se observa en la Ilustración 5-1.

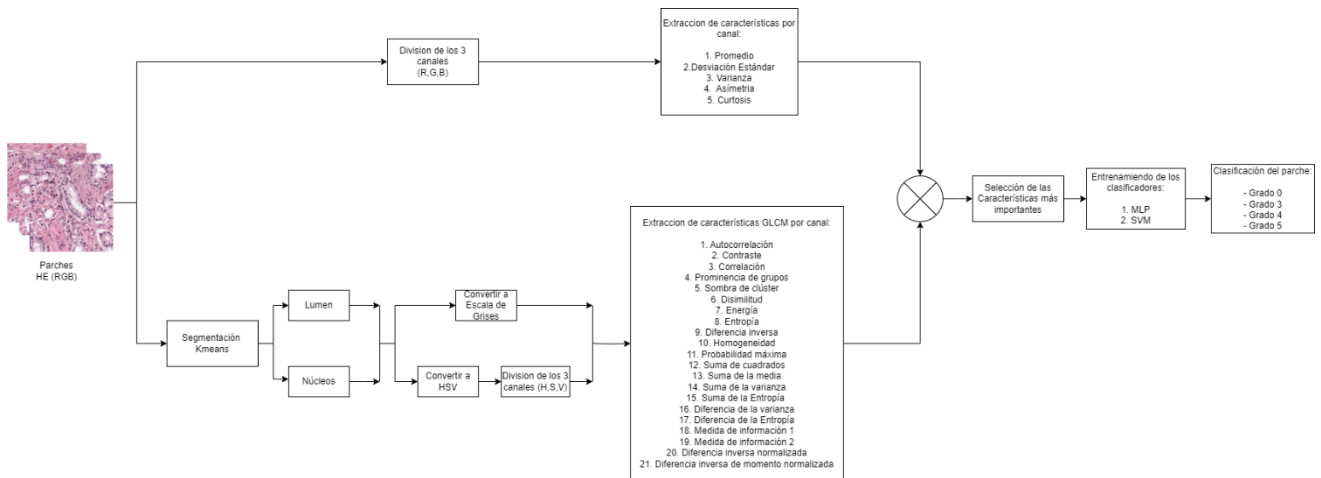


Ilustración 5-1. Solución tomada para la evaluación y extracción de características de las imágenes 512 x 512.

Una vez se hace la unión de las características resulta un conjunto de 183 características, conjunto el cual se somete a una selección de características manual a través de un método embebido por medio de Random Forest con el fin de encontrar el conjunto de características más relevante o que aporte en mayor medida a la clasificación y a su vez aportar robustez y precisión a los clasificadores, característica que no puede ser aportada por medio de relieveF,

pues no tiene en cuenta la reducción de redundancia entre las características, tal y como se menciona en [53].

5.2.5. Resultados y análisis

Al momento de llevar a cabo la prueba, el desempeño en ambos clasificadores aumenta en un 14.5% para MLP y en un 18.9% para SVM en comparación con la prueba 5, confirmando que la unión de características extraídas, el aumento de la dimensión de las imágenes y un método de selección de características en donde no solamente se hace un ranking de las características más relevantes, sino que se tiene en cuenta la importancia de disminuir la redundancia entre características resulta en un mejor desempeño para los clasificadores.

Tabla 5-2. Prueba 2 máquinas superficiales tamaño de imágenes 512x512.

Prueba	Etapas	Características	Clasificadores	Exactitud
Color moments + k means clustering	Preprocesamiento y Extracción de características	183 características	SVM	SVM:0.753
		reducidas a 74 características	MLP	MLP:0.773

Los clasificadores presentan además los siguientes desempeños por clase:

Tabla 5-3. Métricas desempeños máquinas superficiales.

Clasificador	Puntuación	Precisión	Sensibilidad	F1 score
SVM	G0	0.779	0.880	0.826
	G3	0.724	0.760	0.742
	G4	0.725	0.5000	0.592
	G5	0.770	0.870	0.817
	Promedio	0.750	0.753	0.744
MLP	G0	0.793	0.840	0.816
	G3	0.740	0.770	0.755
	G4	0.694	0.680	0.687
	G5	0.870	0.800	0.833
	Promedio	0.774	0.773	0.773

De donde se observa que las mejores puntuaciones clasificadas son la puntuación Gleason 0 para SVM y 5 para MLP, puntuación Gleason 4 es la clase con menor porcentaje de sensibilidad y F1 score en ambas máquinas.

Por otra parte, se destaca que los hiperparámetros con los que se obtuvieron estos resultados son:

Tabla 5-4. Hiperparámetros clasificador SVM.

Kernel	Grado kernel	Parámetro regularización	Estrategia de clasificación
Lineal	3	100	One vs One

Tabla 5-5. Hiperparámetros clasificador MLP.

Número de capas ocultas	Neuronas capa oculta	Función de activación	Optimizador	Parámetro de regularización L2	Número de iteraciones
1	292	ReLU	Adam	$\alpha = 0.00303415$	400

Estos parámetros se obtuvieron a través de funciones de optimización de hiperparámetros, pues al ser un clasificador que tiene muchos hiperparámetros que se pueden configurar, lo mejor es buscar de forma óptima dichos parámetros a través de estas funciones que facilitan el trabajo a diferencia de la forma manual.

5.3. Pruebas – Máquinas profundas

5.3.1. Descripción

En los siguientes apartados sobre las máquinas profundas se encuentra registrado lo realizado con las imágenes de 256x256 y 512x512 píxeles con las máquinas profundas escogidas que fueron 3: AlexNet, ResNet50 y EfficientNetB0 con base en los antecedentes [11], [16] encontrados y al tipo de redes neuronales, que al ser CNN se tienen por redes eficientes para la clasificación de imágenes.

5.3.2. Procedimiento: Imágenes con tamaño de 256x256 píxeles

En el caso de las máquinas de aprendizaje profundo, no es esencial implementar técnicas de preprocesamiento o segmentación de manera explícita por parte del investigador. Esto se debe a que estas máquinas tienen la capacidad de identificar y extraer de manera independiente los descriptores o características significativas directamente del conjunto de datos de entrada. De esta manera, se entrega un conjunto de imágenes acotadas en las dos dimensiones siendo primero las imágenes de 256x256 píxeles entregadas a las 3 redes profundas.

Por otro lado, se llevó a cabo transfer learning por lo que no fue necesario un entrenamiento desde cero en las redes profundas, es decir, se toman las 3 redes previamente entrenadas en problemas de clasificación de imágenes previos, con el fin actualizar sus respectivos pesos al problema de clasificación del adenocarcinoma de próstata. Esto se hizo para evitar el sobre ajuste de las redes a los datos y evitar sesgo en la clasificación de la puntuación Gleason y para reducir la demanda computacional que conlleva modelos como estos con millones de hiperparámetros para establecer.

5.3.3. Resultados y análisis

Al momento de la clasificación con las imágenes se observa que con un tamaño de 256x256 píxeles los resultados son mejores que en las máquinas superficiales, en donde las redes ResNet50 y EfficientNetB0 presentan los mejores resultados de las 3 redes, tal como se ve en la Tabla 5-6.

Tabla 5-6. Resultados clasificación multiclase máquinas profundas con imágenes 256x256.

Prueba	Optimizador compilación del modelo	Activación capa de salida	Iteraciones	Tasa de aprendizaje	Exactitud
1. AlexNet	Adam	Softmax	10	0.000007	0.679
2. ResNet50	Adam	Softmax	5	0.0000055	0.705
3. EfficientNetB0	Adam	Softmax	12	0.000015	0.746

El desempeño por clase para las diferentes redes fue el siguiente:

Tabla 5-7. Métricas de desempeño por clase.

Clasificador	Puntuación	Precisión	Sensibilidad	F1 score
AlexNet	G0	0.663	0.762	0.709
	G3	0.694	0.685	0.690
	G4	0.610	0.489	0.543
	G5	0.733	0.776	0.754
	Promedio	0.675	0.678	0.674
ResNet50	G0	0.698	0.774	0.735
	G3	0.761	0.691	0.725
	G4	0.633	0.540	0.583
	G5	0.719	0.809	0.762
	Promedio	0.703	0.704	0.701
EfficientNetB0	G0	0.729	0.826	0.774
	G3	0.771	0.745	0.758
	G4	0.685	0.596	0.637
	G5	0.791	0.813	0.802
	Promedio	0.744	0.745	0.743

Teniendo en cuenta la Tabla 5-6, se observa que la mejor red fue EfficientNetB0 con un porcentaje de exactitud de 74.6%, luego ResNet50 con 70.5% y por último AlexNet con 67.9%. Si bien los resultados no representan un diagnóstico cercano o igual al 100% en la clasificación de la severidad del adenocarcinoma de próstata, teniendo en cuenta que se trata de un contexto médico, es necesario que al momento de dar un diagnóstico se revise de manera conjunta con otro especialista de apoyo. Por tanto, resulta importante la revisión de la clasificación de la puntuación Gleason 4, pues presenta en términos de métricas de evaluación los resultado con menor alcance.

Con base en esto la Tabla 5-7 se decide probar con el conjunto de imágenes de 512x512 pixeles para observar si los resultados obtenidos con los parches de las imágenes de 256x256 pueden ser mejorados.

5.3.4. Procedimiento: Imágenes con tamaño de 512x512 pixeles

Después de probar las 3 arquitecturas y ver los resultados lo que se hizo fue cambiar el conjunto de imágenes por el de 512x512 pixeles y ver si los resultados presentan mejoría.

5.3.5. Resultados y análisis

El análisis de los resultados (Tabla 5-8) evidencia un aumento de aproximadamente 12% en el desempeño de la clasificación de las diferentes puntuaciones Gleason:

Tabla 5-8. Resultados máquinas profundas imágenes de 512x512 pixeles.

Prueba	Optimizador compilación del modelo	Activación capa de salida	Iteraciones	Tasa de aprendizaje	Exactitud
1. AlexNet	Adam	Softmax	50	0.000007	0.735
2. ResNet50	Adam	Softmax	5	0.0000055	0.859
3. EfficientNetB0	Adam	Softmax	12	0.000015	0.844

Además, los demás clasificadores presentan las siguientes métricas demás:

Tabla 5-9. Métricas de desempeños redes profundas.

Clasificador	Puntuación	Precisión	Sensibilidad	F1 score
AlexNet	G0	0.693	0.879	0.775
	G3	0.757	0.667	0.709
	G4	0.719	0.585	0.645
	G5	0.779	0.810	0.794
	Promedio	0.737	0.735	0.731
ResNet50	G0	0.819	0.939	0.875
	G3	0.901	0.854	0.877
	G4	0.877	0.766	0.818
	G5	0.849	0.878	0.863
	Promedio	0.862	0.859	0.858
EfficientNetB0	G0	0.848	0.899	0.873
	G3	0.856	0.833	0.844
	G4	0.824	0.751	0.786
	G5	0.847	0.893	0.869
	Promedio	0.844	0.844	0.843

Al aumentar el tamaño de las imágenes de 256x256 a 512x512 pixeles con las máquinas profundas se logra también el aumento en el desempeño de las redes profundas en un 6%

para AlexNet y más del 10% para ResNet50 y para EfficientNetB0, en donde se nota los mejores desempeños entre la red ResNet50 y EfficientNetB0.

Las puntuaciones 5 y 0 obtuvieron mejor clasificación para EfficientNetB0 y ResNet50 respectivamente y la puntuación 4 obtuvo la menor tasa de acierto en la clasificación en ambas máquinas.

5.4. Diagnóstico por tejidos – Análisis Comparativo

5.4.1. Descripción

En este apartado se evalúan las distintas máquinas SVM, MLP, ResNet50 y EfficientNetB0, las cuales fueron previamente entrenadas para diagnosticar la puntuación Gleason del parche. De tal modo, que el objetivo ya no es la clasificación de un conjunto de parches, sino el diagnóstico de un tejido completo referente a un paciente, en donde se inicia con el proceso de obtención de parches para diagnosticar cada parche y al final dar un veredicto sobre el tejido que se quiere evaluar. La prueba se llevó a cabo evaluando un conjunto de 120 imágenes de tejidos los cuales los clasificadores no usaron para entrenar y clasificar (Ver Ilustración 4-2).

5.4.2. Procedimiento

Esta prueba funciona como un sistema al cual se le entrega una imagen microscópica de una biopsia del tejido prostático perteneciente a un grupo de imágenes nuevas o un conjunto nuevo de imágenes que los clasificadores se encargarán de clasificar y diagnosticar el grado de severidad del tejido. El sistema se encarga de realizar la división en parches de las imágenes de tamaño de 512x512, por ventanas deslizantes, y realiza los diagnósticos de la severidad del cáncer para todo el tejido con base en los clasificadores con los mejores desempeños de las máquinas superficiales y profundas que son SVM, MLP, EfficientNetB0 y ResNet50.

Una vez son realizadas las clasificaciones de los parches extraídos el proceso que sigue es dar el diagnóstico final del grado con base en las distintas clasificaciones de las diferentes puntuaciones detectadas por las máquinas en los tejidos (Ilustración 4-2). A medida que los clasificadores hacen las predicciones de los parches el sistema va almacenando cuantos parches se clasifican en la escala de Gleason (0, 3, 4 y 5) y con base en la cantidad total de parches extraídos que fueron evaluados, se saca el porcentaje de cada puntuación de Gleason presente en el tejido. De este manera, en función de esos porcentajes y cantidad de parches por grado, presentes dentro del tejido, se toma la decisión final de la severidad definitiva del tejido que se está evaluando en una escala de 0 a 5, a través de una MLP.

5.4.3. Resultados y análisis

Al llevar a cabo se obtuvo los siguientes resultados para los 4 clasificadores como se observa en la Tabla 5-10:

Tabla 5-10. Resultados evaluación del tejido completo.

Clasificadores	Exactitud
SVM	0.542
MLP	0.536
EfficientNetB0	0.957
ResNet50	0.954

Por clase además se obtuvo las siguientes métricas:

Tabla 5-11. Métricas por clase clasificación tejido completo.

Clasificador	Grado	Precisión	Sensibilidad	F1 score
SVM	G0	0.781	0.864	0.820
	G1	0.583	0.650	0.615
	G2	0.349	0.371	0.360
	G3	0.390	0.329	0.357
	G4	0.452	0.407	0.429
	G5	0.647	0.629	0.638
	Promedio	0.553	0.542	0.536
MLP	G0	0.685	0.807	0.741
	G1	0.520	0.571	0.544
	G2	0.426	0.429	0.427
	G3	0.277	0.164	0.206
	G4	0.531	0.550	0.540
	G5	0.638	0.693	0.664
	Promedio	0.513	0.536	0.521
EfficientNetB0	G0	0.993	0.993	0.993
	G1	0.957	0.957	0.957
	G2	0.943	0.950	0.947
	G3	0.904	0.943	0.923
	G4	0.963	0.921	0.942
	G5	0.986	0.979	0.982
	Promedio	0.958	0.957	0.957
ResNet50	G0	0.986	0.993	0.989
	G1	0.972	0.986	0.979
	G2	0.942	0.929	0.935
	G3	0.897	0.936	0.916
	G4	0.941	0.910	0.924
	G5	0.986	0.971	0.978
	Promedio	0.954	0.954	0.954

De los resultados obtenidos se ve una disminución de los desempeños de las máquinas superficiales con base en los resultados obtenidos de la clasificación multiclase de solo

parches, por otra parte, en las redes profundas ResNet50 y EfficientNetB0 lograron ajustarse mejor a la configuración hecha para la clasificación alcanzando desempeños de 95.36% y 95.71% respectivamente.

Los hiperpárametros establecidos para los diferentes MLP usadas en los 4 clasificadores son los que se registran en la Tabla 5-12.

Tabla 5-12. Hiperpárametros máquinas MLP diagnóstico tejido completo.

Clasificador	Número de capas ocultas	Neuronas capa oculta	Función de activación	Optimizador	Parámetro de regularización L2	Número de iteraciones
SVM	1	120	ReLU	Adam	$\alpha = 0.1$	200
MLP	1	120	ReLU	Adam	$\alpha = 0.1$	200
ResNet50	1	120	ReLU	Lbfgs	$\alpha = 0.06$	200
EfficientNetB0	1	120	ReLU	Lbfgs	$\alpha = 0.00000003$	200

En la Tabla 5-12, se registran los hiperpárametros configurados para las distintas máquinas MLP usadas para el diagnóstico final del tejido en función de la clasificación por parches los 4 clasificadores a través optimización de hiperpárametros.

5.5. Alcances

Las máquinas superficiales después del desarrollo y las diferentes pruebas implementadas son las que logran los mayores resultados de los 4 clasificadores superficiales probados con 0.754 y 0.773 para SVM y MLP respectivamente, estos resultados son cercanos a lo reportado en los antecedentes consultados para una clasificación multiclase por parches, teniendo en cuenta la división por puntuación Gleason 0, 3 4 y 5. Se debe tener en cuenta que la base de datos usada es diferente de la revisión bibliográfica.

Con las máquinas profundas al igual que con las máquinas superficiales se logra una mejora en los resultados de clasificación llegando hasta 0.859 y 0.844 para ResNet50 y EfficientNetB0 que son también resultados cercanos a los trabajos previos en antecedentes donde se usaron diferentes tipos de máquinas de aprendizaje profundo [11], [16], para una clasificación multiclase por parche entre las 4 clases entre las puntuaciones Gleason 0, 3, 4 y 5.

El máximo desempeño en este trabajo se alcanza con la evaluación del tejido completo, en donde las redes EfficientNetB0 y ResNet50, tuvieron resultado de 0.957 y 0.954, respectivamente, logrando discriminar con un nivel alto de confianza las diferentes severidades del tejido completo en una escala de 0 a 5. En el caso de las máquinas superficiales los desempeños se redujeron a 0.542 y 0.536 para SVM y MLP respectivamente, debido a que la capacidad de las características con que se entrenaron estos clasificadores no fueron capaces de otorgar la facultad de generalización para la clasificación de las 120 nuevas imágenes a diagnosticar.

Cabe resaltar que para las máquinas superficiales si bien se variaron las técnicas utilizadas en las etapas de preprocesamiento hay etapas en las que no, así como también el tipo de características extraídas. Para el caso de las máquinas profundas el tipo de entrenamiento fue transfer learning en donde las máquinas han sido previamente entrenadas y probadas para tareas de clasificación distintas y en donde simplemente se hace una actualización de pesos en las neuronas con base en el conjunto de datos con el que se trabajó y así evitar el sobre entrenamiento.

5.6. Limitaciones

Para las máquinas superficiales se encuentra la limitación en la capacidad de lograr desempeños de clasificación fiables o aumento de este a través de la clasificación de todo el tejido, pues, aunque se logró 0.754 y 0.773 de exactitud en la clasificación multiclase para SVM y MLP, en la evaluación completa y diagnóstico por tejido el desempeño se redujo considerablemente teniendo un desempeño de 0.542 y 0.536. De igual forma se resalta que si bien se hicieron varias pruebas y se varió en la aplicación de ciertas técnicas, hubo partes del proceso que no se variaron, como por ejemplo el tipo de características extraídas, pues si bien se extrajeron características de textura y de contorno, dentro de estas características no se varió el tipo de obtención de estas características, pues siempre fueron de textura, con la obtención de la matriz GLCM y características estadísticas por canales del modelo R,G y B y por contorno fue de tipo morfológicas con una sola aplicación de las 5 pruebas hechas.

En la etapa de preprocesamiento si bien se reduce el ruido presente en las imágenes debido a la tinción que se hace con colorantes químicos, a través de la aplicación de restricciones de valores de pixeles correspondientes a los colores diferentes al tejido por tinción, dicho ruido seguía presente en las imágenes solo que en menor cantidad y si afectando en menor cantidad que un principio, pero de la misma forma afectando el proceso de obtención de características y posterior clasificación.

En la etapa de segmentación mediante el uso de la técnica de Watershed se tuvo la limitante de vectores de características pobres debido a la segmentación, pues se logra segmentar de forma satisfactoria según los índices obtenidos en la Tabla 4-2, pero según los resultados al hacer la clasificación con las características de tipo morfológicas los desempeños no alcanzaron a superar como mínimo más del 50% de exactitud. También se identifica que la presencia de ruido afecta el proceso de segmentación al momento de hacer la distinción entre las células y el resto de la imagen.

En las máquinas profundas la limitación en los desempeños alcanzados se debe, en parte, a la falta de variaciones en el proceso de entrenamiento, el cual fue hecho a través de transfer learning, lo que resulta si bien en una actualización de los pesos al conjunto de datos a usar pero que difiere a si se hubiesen entrenado desde cero, pero que desde luego implica una demanda mayor en cuanto al recurso computacional que es otra limitante constante al momento de procesar una cantidad de datos alta y que no asegura un modelo con generalización en el aprendizaje debido al conjunto de imágenes disponibles a usar, por ello

el entrenamiento se limitó a una actualización de hiperparámetros de modelos pre entrenados por transfer learning.

En adición al tema del recurso computacional otra limitante fue la adquisición y la capacidad de poder trabajar con más datos, pues el centro Radboud ofrece más de 5160 imágenes de tejidos completos, pero al momento de hacer el parcheo de imágenes o “image patching” la capacidad de memoria alcanzaba casi su límite, acotando así el posible uso de las 5160 imágenes de tejidos completos a solo 409 tejidos prostáticos en total, para usar en la extracción de parches, que gracias a la aplicación del image patching se pudo establecer clases de 1000 parches por puntuaciones Gleason en imágenes de 512x512, pero sí de haber utilizado más tejidos se pudiesen tener más parches de entreno para las distintas máquinas probadas.

5.7. Conclusiones

Del trabajo realizado se encuentra que para la complejidad del problema en relación con las clases a distinguir y los datos que se alcanzaron a usar, los clasificadores que presentan menos susceptibilidad al momento de manejar problemas de alta dimensionalidad como MLP y SVM son una buena opción junto con un conjunto de características donde no predomine un solo tipo de características, sino que estas describan desde diferentes tipos de características los diferentes patrones de los tejidos.

Se destaca entonces el uso de técnicas combinadas como una opción para lograr resultados superiores al 70% de exactitud, en vez de solo usar una técnica por etapa, ya sea en preprocesamiento, segmentación o extracción de características, como también el uso de técnicas que ayuden a la generalización de las máquinas superficiales como la validación cruzada. Debido a que los resultados no alcanzan una precisión diagnóstica cercana al 100% en la clasificación de la severidad del adenocarcinoma de próstata, es crucial, dada la naturaleza médica de esta tarea, que el diagnóstico se realice en colaboración con otro especialista, en caso de querer aplicar estas máquinas en la práctica. Es especialmente relevante revisar la clasificación de la puntuación Gleason 4, ya que, según las métricas de evaluación, es donde se observa un alcance más limitado.

La restricción de la aparición de ruido por coloración por tinción resulta apropiada para evitar que se obtengan características atípicas o valores poco apropiados para describir las diferentes clases. El cambio en la aplicación de la técnica de segmentación proporcionó una mejor distinción entre núcleos y tejido, aumentando el desempeño de los distintos clasificadores por medio de la segmentación por k-means clustering o por agrupamiento, a diferencia de solo aplicar la técnica de segmentación de Watershed

El concentrar la extracción de los vectores de características en base a los modelos de color resultó más apropiado que en escala de grises, brindando un mejor aporte que las características extraídas a través de la conversión a escala de grises.

También del trabajo implementado se resalta que el desempeño de la clasificación de parches individuales utilizando las máquinas de aprendizaje profundo ResNet50 y

EfficientNetB0 fue de 85.9% y 84.4% de exactitud, respectivamente. El desempeño de la evaluación de la gravedad del tejido completo (escala de 0 a 5) alcanzó una exactitud del 95.7% y 95.4% para EfficientNetB0 y ResNet50, en este mismo orden. Siendo así, estas CNN como las mejores en cuanto a capacidad de abstracción y obtención de información relevante y capacidad de clasificación con tan solo la aplicación de transfer learning para el caso de análisis de imágenes histopatológicas. En contraste, con artículos como [11] y [16], destacan evaluaciones en la gravedad del cáncer en la escala de 0-5, usando todo el tejido completo sin hacer un análisis por parche, en donde [11] destaca máquinas como InceptionResNet V2 y ResNet101 V2, teniendo resultados del 84.99% y 81.50%, respectivamente. Por otro lado, en [16] se destaca la máquina EfficientNetB7 con una exactitud de 90.90%. De esta manera, se resalta que utilizar un modelo para la clasificación de la severidad del cáncer del tejido completo, a partir de hacer un análisis por parches, puede proporcionar una evaluación más precisa en la severidad del cáncer, destacando así las zonas afectadas en el tejido, y permitiendo diagnósticos más acertados, en comparación a si se hace una evaluación del tejido completo sin hacer un análisis por parche. Cabe resaltar que una evaluación del tejido completo, sin hacer un análisis por parches, puede requerir una reevaluación completa de todo el tejido si el especialista tiene dudas sobre el diagnóstico. En cambio, si se utiliza una evaluación por parches, esta permite una evaluación acompañada con el especialista, indicándole qué zonas y parches son necesarios examinar nuevamente, ahorrándole tiempo a la hora de definir el diagnóstico completo.

En el caso de las máquinas superficiales, se destaca que no tienen la misma capacidad de desempeño como las máquinas profundas, en donde para las máquinas superficiales como SVM y MLP se resalta que el desempeño de la clasificación de parches individuales es de 75.4% y 77.3%, pero para la evaluación de la gravedad tejido completo se obtuvo 54.2% y 53.6%. En comparación, con antecedentes como [13] y [14], solo destacan una clasificación por parches. En el caso de [13], muestra un rendimiento promedio al evaluar cada puntuación Gleason, con una exactitud de 92.7% a través de una MLP que analiza las características de textura y características de color RGB que se extrajeron para un parche de dimensión de 256x256. En el caso de [14], una SVM y MLP analizan las características morfológicas del Lumen y núcleos presentes en cada parche con un tamaño de 512x512 pixeles, presentando un 92% y 82.5% de rendimiento promedio al evaluar cada puntuación Gleason. De esta manera, de acuerdo con nuestro modelo implementado a través de máquinas superficiales por parches, es necesaria una evaluación acompañada por parte de un especialista, en donde este verificará si las puntuaciones asignadas a cada parche corresponden correctamente de acuerdo con la clasificación Gleason.

Por otra parte, es necesario destacar que la obtención de parches con mayor dimensión permitió que las mejoras tuvieran lugar tanto para las máquinas superficiales y profundas en la clasificación por parches, pues esto permitió más información para extraer.

Además, el uso de solapamiento u overlapping para la evaluación del tejido completo permitió primero obtener más imágenes al conservar parte de los parches anteriores junto a

los que se recortan sucesivamente y permitió aumentar los resultados con las redes profundas que incrementaron los resultados de desempeño.

6. Conclusiones generales

De todo el trabajo realizado que comprende desde una búsqueda y análisis de trabajos previos, planteamiento e implementación de máquinas superficiales y profundas para desarrollar una tarea de clasificación multiclase se concluye que:

Se logró detectar la presencia y el grado Gleason en las imágenes microscópicas de biopsias con adenocarcinoma de próstata, a través de 4 clasificadores, 2 máquinas de aprendizaje superficial y 2 máquinas de aprendizaje profundo. En donde, se obtuvieron mayores resultados para las máquinas profundas en la evaluación de tejidos completos con 95.4% y 95.7% para ResNet50 y EfficientNetB0 respectivamente.

La determinación del ajuste y escalamiento adecuado se hizo por medio de la técnica de Image Patching, encontrando que la dimensión de parches de 512x512 pixeles brindaba una información amplia del tejido celular. Cabe destacar, que en la literatura [13], destacaban que dimensiones de 256x256 pixeles, también eran útiles para este tipo de problema, sin embargo al poner a prueba los distintos clasificadores con los parches de 256x256, la diferencia de desempeños en comparación a parches con dimensión de 512x512 estaba entre un 12 y 19 por ciento, esta diferencia se debe ya que con parches de mayor dimensión se logra capturar una porción más extensa del tejido, permitiendo así detectar con más facilidad patrones estructurales en la forma de las células y glándulas. Por otro lado, también se reconoce el uso del solapamiento entre parches, pues con esto se logra una reducción en la pérdida de información entre los bordes de las imágenes, conservando así la información que se podría perder al hacer un image patching sin solapamiento.

En la extracción de características presentes en imágenes microscopias de biopsias, se destacan las características de contorno y textura, siendo muy útiles tal y como resaltan en [13], [14] y [15]. En donde, para el caso de las características de contorno servían para evidenciar los patrones morfológicos en la forma de la célula, y para el caso de las características de textura, nos ayudaban a evidenciar los cambios en las divisiones celulares y cambios en la apariencia estructural de las glándulas.

Limitar el ruido producido por la coloración de la imagen ayuda a evitar que se capturen características anómalas o inapropiadas al describir diferentes categorías. Al cambiar a una técnica de segmentación diferente (de Watershed a k means clustering), la distinción entre núcleos celulares y tejidos se volvió más clara y el rendimiento de múltiples clasificadores mejoró en comparación con el uso exclusivo de la técnica de segmentación Watershed. Además, la extracción de características basadas en modelos de color es más eficiente que usar escala de grises, lo que resultó en mejoras entre un 3 y 15 por ciento en comparación con las características obtenidas al convertir a escala de grises.

Se resalta la importancia de la mezcla de características extraídas, ya que presenta una mejora en el rendimiento de las máquinas superficiales cuando se mezclan características

extraídas de la matriz GLCM con los momentos de color, en comparación con el uso de un solo tipo de características. Se destaca también que la combinación de característica en la división del núcleo y lumen, sumado a la extracción de características de color de la imagen original, aportan un mayor significado en la diferenciación de una clase de otra.

Se reconoce el aporte de clasificadores superficiales que sean capaces de desempeñar tareas de clasificación bajo espacios de características complejos o problemas con alta dimensionalidad, pues el uso de pocas características para la clasificación de la severidad del cáncer de próstata no permitía diferenciar correctamente entre las diferentes clases. Esto se veía reflejado al utilizar clasificadores como Naive Bayes y kNN, en donde obtuvieron desempeños de 45.8% y 55.4% de exactitud, pero para clasificadores como MLP y SVM, alcanzaron desempeños de 62.8% y 56.4%, como se observa en la prueba 5 de la Tabla 5-1.

Para las máquinas superficiales fue necesario aplicar técnicas de visión artificial y reconocimiento de patrones que permitieran la obtención de una región de interés, pues de esta forma se enfoca la extracción de características en los descriptores que permiten el diagnóstico de la agresividad o severidad del adenocarcinoma de próstata, ya sea enfocando regiones de interés en los núcleos, estructura de las glándulas o lumen presentes en los tejidos prostáticos a través de la aplicación de algoritmos como k-means clustering o la obtención de características por medio del análisis individual de canales del modelo RGB y HSV, junto con una selección de características por Random Forest, en donde se seleccionan las características más relevantes y que menos redundancia poseen entre sí.

Al evaluar las dos máquinas de aprendizaje superficial para la clasificación de la severidad del cáncer de próstata, en función de la puntuación Gleason bajo la clasificación multiclase por parches para todo un tejido. Se obtienen bajos desempeños en el caso de las superficiales de 54.2% y 53.6% para SVM y MLP respectivamente, a diferencia de su aplicación para el diagnóstico por parches en vez del tejido completo, se logra desempeños de exactitud de 75.3% para SVM y 77.3% en MLP. Por otro lado, se observa que estos resultados también están por debajo, entre un 5% (caso MLP) y 18% (caso SVM), en comparación con estudios previos citados como [13] y [14], en los que se destacan por una clasificación por parches. En el caso de [13], muestra un rendimiento promedio al evaluar cada puntuación Gleason, con una exactitud de 92.7% a través de una MLP que analiza las características de textura y características de color RGB que se extrajeron para un parche de dimensión de 256x256. En el caso de [14], una SVM y MLP analizan las características morfológicas del Lumen y núcleos presentes en cada parche con un tamaño de 512x512 píxeles, presentando un 92% y 82.5% de rendimiento promedio al evaluar cada puntuación Gleason. De esta manera, se concluye que para nuestro modelo implementado a través de máquinas superficiales por parches, es necesaria una evaluación acompañada por parte de un especialista, en donde este verificará si las puntuaciones asignadas a cada parche corresponden correctamente de acuerdo con la clasificación Gleason.

En el caso de las máquinas profundas tal como se mencionó anteriormente para la clasificación de la puntuación Gleason a través de los parches de tamaño 512x512 píxeles,

los resultados fueron de 85.9% y 84.4% para ResNet50 y EfficientNetB0 respectivamente, aumentando aún más en la clasificación de tejido completo en 95.4% y 95.7% para ResNet50 y EfficientNetB0 nuevamente. Por otro lado, se observa que estos resultados están por encima, entre un 5% y 11%, en comparación con estudios previos citados como [11] y [16], en los que se destacan por hacer evaluación completa del tejido, en donde [11] destaca máquinas como InceptionResNet V2 y ResNet101 V2, teniendo resultados del 84.99% y 81.50%, respectivamente. En el caso de [16] se destaca la máquina EfficientNetB7 con una exactitud de 90.90%.

Se resalta que las máquinas de aprendizaje profundo implementadas alcanzaron un desempeño entre un 40% y 42% de exactitud por encima de las máquinas superficiales, al realizar una evaluación completa del tejido a través de parches, lo que confirma lo comúnmente aceptado en el campo del análisis de imágenes sobre la superioridad de las redes profundas para tareas de clasificación en imágenes médicas. Esto se debe ya que las redes profundas tienen la capacidad de aprender automáticamente características complejas y abstraer diferentes características de las capas convolucionales, lo que les otorga una mayor capacidad para procesar y abstraer patrones relevantes. Sin embargo, es importante destacar que las máquinas de aprendizaje profundo tienen limitaciones en su aplicación, ya que requieren recursos computacionales significativos. Dependiendo de la cantidad de capas convolucionales, puede que una clasificación tarde en procesarse y en dar su diagnóstico final. Por otro lado, es necesario destacar que el desempeño de estas máquinas podría mejorarse con la disponibilidad de más imágenes y recursos computacionales que permitan un entrenamiento y ajuste de parámetros desde cero, específicamente para el diagnóstico de la severidad del cáncer de próstata

Se resalta que las máquinas de aprendizaje profundo implementadas alcanzaron entre un 40% y 42% de exactitud mayor respecto a las máquinas superficiales, al realizar una evaluación completa del tejido a través de parches, lo que confirma lo comúnmente aceptado en el campo del análisis de imágenes sobre la superioridad de las redes profundas para tareas de clasificación en imágenes médicas. Esto se debe a que las redes profundas tienen la capacidad de aprender automáticamente características complejas y abstraer diferentes características de las capas convolucionales, lo que les otorga una mayor capacidad para procesar y abstraer patrones relevantes. Sin embargo, es importante destacar que las máquinas de aprendizaje profundo tienen limitaciones en su aplicación, debido a que requieren una carga alta de recursos computacionales y tiempo significativo. Dependiendo de la cantidad de capas convolucionales, puede que una clasificación tarde en procesarse y en dar su diagnóstico final.

Si bien el objetivo de este trabajo no es reemplazar la labor del especialista en el trato y diagnóstico al paciente, la intención es por medio de la aplicación de aprendizaje automático y de recursos tecnológicos que permitan un apoyo en este tipo de decisiones, que por más experiencia que se pueda disponer, es posible la duda y desacuerdo por saber qué diagnóstico dar al paciente, por lo que se considera que sistemas llevados a cabo por redes neuronales convolucionales con resultados como el logrado en este trabajo, representan

confiabilidad para los profesionales, pero se reconoce que la decisión del diagnóstico es crucial para el tratamiento de la condición de salud que representa el adenocarcinoma de próstata, por lo que se considera como apto para usar como un apoyo o consulta rápida para los especialistas. Por otro lado, es necesario destacar que el desempeño de estas máquinas de aprendizaje profundo podría mejorarse con la disponibilidad de más imágenes y recursos computacionales que permitan un entrenamiento y ajuste de parámetros desde cero, específicamente para el diagnóstico de la severidad del cáncer de próstata.

Se logra establecer una clasificación robusta frente a la aparición o presencia de características que pueden generar desviaciones al momento de la diferenciación de las clases como lo es el fondo blanco y las coloraciones diferentes por tinción, pues a comparación de las investigaciones [13], [14] y [15] sus métodos de preprocesamiento apuntan a una obtención de parches que, solo concentren información de tejido, en nuestro caso, se permitió que los clasificadores entrenaran y aprendieran las características esenciales sin descartar el fondo o tinciones teniendo en cuenta que se ajustó la selección a parches que concentraran poca información respecto a la tinción presente en los tejidos.

De esta forma, este estudio representa un avance en la intención de generar una clasificación precisa de la severidad del cáncer de próstata mediante el uso de técnicas de aprendizaje profundo y superficial. Se demuestra que, a pesar de los desafíos presentados por variaciones en la tinción y el fondo de las imágenes, las metodologías llevadas a cabo han sido capaces de adaptarse y aprender de estas características en lugar de descartarlas, ofreciendo una visión más integral y robusta del tejido prostático.

Los resultados obtenidos con ResNet50 y EfficientNetB0, especialmente en la clasificación de tejidos completos, son notables, subrayando la superioridad de las redes de aprendizaje profundo en tareas complejas de clasificación de imágenes. Además, la implementación de técnicas como el image patching con solapamiento y el transfer learning ha demostrado ser fundamental en la mejora del rendimiento. Este trabajo no solo representa un salto cualitativo en el diagnóstico del cáncer de próstata, dentro de las investigaciones y trabajos hechos previamente, sino que también establece un precedente para futuras investigaciones y aplicaciones clínicas, abriendo el camino para diagnósticos más precisos, tempranos y personalizados. Por lo tanto, nuestra investigación subraya la importancia crítica de la integración continua de innovaciones tecnológicas en el campo de la medicina diagnóstica, prometiendo un futuro donde la inteligencia artificial y la patología digital juegan un papel central en la mejora de los resultados de salud de los pacientes.

7. Trabajos futuros

Como trabajos futuros para las máquinas superficiales se tiene como propuesta mejorar el proceso de image patching o la obtención de parches de menor dimensión a la original pues a pesar de que se aplicaron condiciones de color en base a la media y desviación estándar de los píxeles en las imágenes para seleccionar imágenes con menos cantidad de fondo blanco y aquellas que se veían afectar por la tinción con los químicos (imágenes con coloración azul, café, gris y negro), quedaban aun así imágenes afectadas por este tipo de

ruido, que si bien se redujo considerablemente, su presencia adiciona valores como ruido que afectan la clasificación. El uso de un conjunto de características en base a la unión de distintos tipos de características. También la posible opción de extracción de características por medio de redes neuronales con el fin de aprovechar su capacidad de abstracción y usar esas características para entrenar los clasificadores. Realizar más tipos de extracción de características de textura y estadísticas.

Además, respecto a la selección del conjunto de características se recomienda como trabajo futuro usar algoritmos de selección de características que ajusten dichos conjuntos a las máquinas a usar, pues si bien se utilizó el método embebido para Random Forest y con esto se logró un aumento en la clasificación para SVM y MLP, sería más apropiado hacer lo mismo para que se ajuste a SVM y MLP o a los clasificadores que se quieren probar.

Para las máquinas profundas un primer trabajo futuro es aumentar los datos de entrenamiento y prueba, esto también para máquinas superficiales, pero aún más para las máquinas profundas pues si se quiere retomar las máquinas profundas usadas en este trabajo, son redes con millones de parámetros para entrenar. Además, entrenar estas máquinas desde cero con solo la base de datos de este problema, con el fin de que el aprendizaje se ajuste más al tipo de tejidos de cáncer de próstata, obviamente manejando más datos de los disponibles de la base de datos pues esto requerirá mayor cantidad de datos. Otra opción para considerar es el uso de técnicas de optimización de hiperparámetros con el fin de mejorar los desempeños con este tipo de máquinas profundas y para las máquinas superficiales una buena opción a considerar son las etapas de post procesamiento que ayuden a aumentar los desempeños aumentos que para este trabajo no se consideraron.

Para trabajar en la mejora de resultados de clasificación también se recalca la importancia del manejo de diferentes datasets, para evaluar y mejorar la capacidad de generalización de los clasificadores, evaluar nuevas arquitecturas junto a modelos de aprendizaje profundo, ya sea por medio de variación de hiperparámetros o conjuntos de redes neuronales convolucionales, que se destacan por su destreza en la clasificación en estas aplicaciones, evaluar y contrastar los resultados con profesionales de la salud, especialistas en el diagnóstico del adenocarcinoma de próstata, en consejos o posibles ideas como abordar el problema.

8. Referencias bibliográficas

- [1] MedlinePlus, “Cáncer de próstata”, Biblioteca Nacional de Medicina, EE.UU, 2022. Consultado: el 12 de mayo de 2022. [En línea]. Disponible en: <https://medlineplus.gov/spanish/prostatecancer.html>
- [2] Cancer.Net, “Cáncer de Próstata: Estadísticas”, American Society of Clinical Oncology, EE.UU, 2022. Consultado: el 12 de mayo de 2022. [En línea]. Disponible en: <https://www.cancer.net/es/tipos-de-c%C3%A1ncer/c%C3%A1ncer-de-pr%C3%B3stata/estad%C3%ADsticas>
- [3] Ministerio de salud, “Minsalud se unió a conmemoración del Día Mundial contra el Cáncer”, Bogotá, Colombia, 2022. Consultado: el 12 de mayo de 2022 [En línea]. Disponible en: <https://www.minsalud.gov.co/Paginas/Minsalud-se-unio-a-conmemoraci%C3%B3n-del-Dia-Mundial-contra-el-Cancer.aspx>
- [4] MedlinePlus, “Sistema de puntuación de Gleason”, Biblioteca Nacional de Medicina, EE.UU, 2022. Consultado: el 12 de mayo de 2022. [En línea]. Disponible en: <https://medlineplus.gov/spanish/ency/patientinstructions/000920.htm>
- [5] J. Persson *et al.*, “Interobserver variability in the pathological assessment of radical prostatectomy specimens: findings of the Laparoscopic Prostatectomy Robot Open (LAPPRO) study”, *Scand J Urol*, vol. 48, núm. 2, pp. 160–167, 2014. [En línea]. Disponible doi: 10.3109/21681805.2013.820788.
- [6] R. Montironi, R. Mazzuccheli, M. Scarpelli, A. Lopez-Beltran, G. Fellegara, y F. Algaba, “Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: Contemporary approach, current clinical significance and sources of pathology discrepancies”, *BJU Int*, vol. 95, núm. 8, pp. 1146–1152, jun. 2005. [En línea]. Disponible doi: 10.1111/j.1464-410X.2005.05540.x.
- [7] Q. Li, R.M. Nishikawa. “Computer-Aided Detection and Diagnosis in Medical Imaging”, 1st Edition, Boca Raton: CRC Press, 2015. [En línea]. Disponible doi: 10.1201/b18191
- [8] W. Bulten, K. Kartasalo, PH. Chen. et al. “Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge”, *Nat Med* 28, 154–163, 2022. [En línea]. Disponible doi: 10.1038/s41591-021- 01620-2
- [9] Y. Gutiérrez, G. Garzón, y F. Martínez, “Towards clinical significance prediction using ktrans evidences in prostate cancer”, *2019 22nd Symposium on Image, Signal Processing and Artificial Vision, STSIVA 2019 - Conference Proceedings*, abr. 2019. [En línea]. Disponible doi: 10.1109/STSIVA.2019.8730282
- [10] H. Kallen, J. Molin, A. Heyden, C. Lundstrom, y K. Astrom, “Towards grading gleason score using generically trained deep convolutional neural networks”, *Proceedings -*

- International Symposium on Biomedical Imaging, vol. 13, pp. 1163–1167, jun. 2016. [En línea]. Disponible doi: 10.1109/ISBI.2016.7493473.
- [11] G. P. Kanna, S. J. K. J. Kumar, P. Parthasarathi, y Y. Kumar, “A Review on Prediction and Prognosis of the Prostate Cancer and Gleason Grading of Prostatic Carcinoma Using Deep Transfer Learning Based Approaches”, *Archives of Computational Methods in Engineering*, vol. 30, núm. 5, pp. 3113–3132, jun. 2023. [En línea]. Disponible doi: 10.1007/S11831-023-09896-Y/METRICS.
 - [12] H. Xu, S. Park, y T. H. Hwang, “Computerized Classification of Prostate Cancer Gleason Scores from Whole Slide Images”, *IEEE/ACM Transactions Computational Biology and Bioinformatics*, vol. 17, núm. 6, pp. 1871–1882, nov. 2020. [En línea]. Disponible doi: 10.1109/TCBB.2019.2941195.
 - [13] S. Bhattacharjee et al., “Multi-Features Classification of Prostate Carcinoma Observed in Histological Sections: Analysis of Wavelet-Based Texture and Colour Features”, *Cancers* 2019, vol. 11, núm. 12, p. 1937, dic. 2019. [En línea]. Disponible doi: 10.3390/CANCERS11121937.
 - [14] S. Bhattacharjee *et al.*, “Quantitative Analysis of Benign and Malignant Tumors in Histopathology: Predicting Prostate Cancer Grading Using SVM”, *Applied Sciences* 2019, vol. 9, núm. 15, p. 2969, jul. 2019. [En línea]. Disponible doi: 10.3390/APP9152969.
 - [15] D. Albashish, S. Sahran, A. Abdullah, N.A. Shukor, S. Hayati, Md. Pauzi. “Lumen-Nuclei Ensemble Machine Learning System for Diagnosing Prostate Cancer in Histopathology Images”, *Pertanika Journal of Science and Technology*, vol. 25. pp. 39 – 48, 2017. [En línea]. Disponible en: https://www.researchgate.net/publication/324687370_Lumen-Nuclei_Ensemble_Machine_Learning_System_for_Diagnosing_Prostate_Cancer_in_Histopathology_Images
 - [16] M. Gour, S. Jain, y U. Shankar, “Application of Deep Learning Techniques for Prostate Cancer Grading Using Histopathological Images”, *Communications in Computer and Information Science*, vol. 1567 CCIS, pp. 83–94, 2022. [En línea]. Disponible doi: 10.1007/978-3-031-11346-8_8/COVER.
 - [17] Sociedad Americana del Cáncer, “¿ Qué es el cáncer de próstata?”, 2019. Consultado: el 12 de mayo de 2022. [En línea]. Disponible en: <https://www.cancer.org/es/cancer/tipos/cancer-de-prostata/acerca/que-es-cancer-de-prostata.html>
 - [18] Instituto Nacional del Cáncer, “Adenocarcinoma”, *Diccionario del Cáncer*. Consultado: el 21 de febrero de 2024. [En línea]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/adenocarcinoma>

- [19] Instituto Nacional del Cáncer, "Tratamiento del cáncer de próstata", Tratamientos del Cáncer. Consultado: el 21 de febrero de 2024. [En línea]. Disponible en: <https://www.cancer.gov/espanol/tipos/prostata/paciente/tratamiento-prostata-pdq>
- [20] Instituto Nacional del Cáncer, "Biopsia", Diccionario del Cáncer. Consultado: el 12 de mayo de 2022. [En línea]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/biopsia>
- [21] Instituto Nacional del Cáncer, "Tinción con hematoxilina y eosina", Diccionario del Cáncer. Consultado: el 21 de febrero de 2024. [En línea]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/tincion-con-hematoxilina-y-eosina>
- [22] Sociedad Americana del Cáncer, "Pruebas para diagnosticar y determinar la etapa del cáncer de próstata", 2022. Consultado: el 12 de mayo de 2022. [En línea]. Disponible en: <https://www.cancer.org/es/cancer/tipos/cancer-de-prostata/deteccion-diagnostico-clasificacion-por-etapas/como-se-diagnostica.html>
- [23] N. Chen y Q. Zhou, "The evolving Gleason grading system", *Chinese Journal of Cancer Research*, vol. 28, núm. 1, p. 58, feb. 2016. [En línea]. Disponible doi: 10.3978/J.ISSN.1000-9604.2016.02.04.
- [24] C. A. Gallardo, "Micrografías electrónicas: Una herramienta para la conservación preventiva de materiales fotográficos", *Alquimia, Sistema Nacional de Fototecas: Salud pública*, núm. 73, pp. 81-82, mayo de 2022. [En línea]. Disponible en: <https://www.revistas.inah.gob.mx/index.php/alquimia/issue/view/2591/2820>
- [25] H. Fu, G. Qiu, J. Shu, and M. Ilyas, "A novel polar space random field model for the detection of glandular structures," *IEEE Trans Med Imaging*, vol. 33, no. 3, pp. 764–776, 2014. [En línea]. Disponible doi: 10.1109/TMI.2013.2296572.
- [26] Scikit-learn, "6.3. Preprocessing data", Scikit-learn Documentation, versión 1.3.2. Consultado: el 14 de noviembre de 2023. [En línea]. Disponible en: <https://scikit-learn.org/stable/modules/preprocessing.html>
- [27] J. Singnoo, G. D. Finlayson. "Understanding the Gamma Adjustment of Images", *Proc. IS&T 18th Color and Imaging Conference*, pp. 134–139, ene. 2010. [En línea]. Disponible doi: 10.2352/CIC.2010.18.1.art00024
- [28] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, núm. 7, pp. 674–693, 1989. [En línea]. Disponible doi: 10.1109/34.192463.
- [29] A. D. Belsare, "Histopathological Image Analysis Using Image Processing Techniques: An Overview", *Signal Image Process*, vol. 3, núm. 4, pp. 23–36, ago. 2012. [En línea]. Disponible doi: 10.5121/SIPIJ.2012.3403.

- [30] Ş. Öztürk y B. Akdemir, "Cell-type based semantic segmentation of histopathological images using deep convolutional neural networks", *Int. J. Imaging Systems and Technology*, vol. 29, núm. 3, pp. 234–246, sep. 2019. [En línea]. Disponible doi: 10.1002/IMA.22309.
- [31] X. Xu, S. Xu, L. Jin, y E. Song, "Characteristic analysis of Otsu threshold and its applications", *Pattern Recognition Letters*, vol. 32, núm. 7, pp. 956–961, may 2011. [En línea]. Disponible doi: 10.1016/J.PATREC.2011.01.021.
- [32] X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, Y. Zheng, "Learning Active Contour Models for Medical Image Segmentation", *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11624-11632, 2019. [En línea]. Disponible doi: 10.1109/CVPR.2019.01190
- [33] G. Hamarneh y X. Li, "Watershed segmentation using prior shape and appearance knowledge", *Image and Vision Computing*, vol. 27, núm. 1–2, pp. 59–68, ene. 2009. [En línea]. Disponible doi: 10.1016/J.IMAVIS.2006.10.009.
- [34] Y. Li y H. Wu, "A Clustering Method Based on K-Means Algorithm", *Physics Procedia*, vol. 25, pp. 1104–1109, ene. 2012. [En línea]. Disponible doi: 10.1016/J.PHPRO.2012.03.206.
- [35] M. Singh, E. M. Kalaw, D. M. Giron, K.-T. Chong, C. L. Tan, y H. K. Lee, "Gland segmentation in prostate histopathological images", *Journal of Medical Imaging*, vol. 4, núm. 2, p. 027501, jun. 2017. [En línea]. Disponible doi: 10.1117/1.JMI.4.2.027501.
- [36] S. Ozdemir y D. Susarla, *Feature Engineering Made Easy: Identify Unique Features from Your Dataset in Order to Build Powerful Machine Learning Systems*. Birmingham, UK: Packt Publishing, 2018.
- [37] R. M. Haralick, I. Dinstein, y K. Shanmugam, "Textural Features for Image Classification", *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, núm. 6, pp. 610–621, 1973. [En línea]. Disponible doi: 10.1109/TSMC.1973.4309314.
- [38] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions", *SN Comput Sci*, vol. 2, núm. 3, pp. 1–21, may 2021. [En línea]. Disponible doi: 10.1007/S42979-021-00592-X/FIGURES/11.
- [39] ScienceDirect, "Pattern Recognition", Elsevier, 2023. Consultado: el 14 de noviembre de 2023. [En línea]. Disponible en: <https://www.sciencedirect.com/journal/pattern-recognition>
- [40] H. Liu y H. Motoda, "Computational Methods of Feature Selection", *Computational Methods of Feature Selection*, pp. 1–419, oct. 2007. [En línea]. Disponible doi: 10.1201/9781584888796

- [41] S. Sah, "Machine Learning: A Review of Learning Types", Preprints, jul. 2020. [En línea]. Disponible doi: 10.20944/PREPRINTS202007.0230.V1.
- [42] S. Ranganathan, K. Nakai y C. Schonbach, "Encyclopedia of Bioinformatics and Computational Biology", ABC of Bioinformatics, 1ra ed. Amsterdam, Netherlands: Elsevier, 2018, pp. 403-411.
- [43] P. Cunningham y S. J. Delany, "k-Nearest Neighbour Classifiers-A Tutorial", ACM Comput. Survey, vol. 54, 2021. [En línea]. Disponible doi: 10.1145/3459665.
- [44] D. A. Pisner y D. M. Schnyer, "Support vector machine", *Machine Learning: Methods and Applications to Brain Disorders*, pp. 101–121, ene. 2020. [En línea]. Disponible doi: 10.1016/B978-0-12-815739-8.00006-7.
- [45] A. Singh, N. Thakur, A. Sharma, "A review of supervised machine learning algorithms", *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2016, pp. 1310-1315.
- [46] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems*, vol. 25, 2012. [En línea]. Disponible en: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [47] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770-778. [En línea]. Disponible doi: 10.1109/CVPR.2016.90
- [48] M. Tan y Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", *International Conference on Machine Learning*, 2019. [En línea]. Disponible doi:10.48550/arXiv.1905.11946
- [49] D. M. W. Powers y Ailab, "Evaluation: from precision, recall and F-measure to ROC, Informedness, Markedness and Correlation", *International Journal of Machine Learning Technology*, vol. 2, pp.37-63, 2011. [En línea]. Disponible doi: 10.48550/arXiv.2010.16061
- [50] M. Macenko et al., "A method for normalizing histology slides for quantitative analysis", *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Boston, MA, USA, 2009, pp. 1107-1110. [En línea]. Disponible doi: 10.1109/ISBI.2009.5193250.
- [51] M. A. Morshidi, M. H. Marhaban, y A. Jantan, "Color segmentation using multi layer neural network and the HSV color space", *Proceedings of the International Conference on Computer and Communication Engineering 2008, ICCCE08: Global Links for Human Development*, pp. 1335–1339, 2008. [En línea]. Disponible doi: 10.1109/ICCCE.2008.4580822.

- [52] D. Teng, F. Li, y W. Zhang, “Using comprehensive machine-learning models to classify complex morphological characters”, *Ecology and Evolution*, vol. 11, núm. 15, pp. 10421–10431, ago. 2021. [En línea]. Disponible doi: 10.1002/ECE3.7845.
- [53] R. Revathy y R. Lawrance, “Classifying crop pest data using C4.5 algorithm”, *Proceedings of the 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS 2017*, vol. 2018-February, pp. 1–6, jul. 2017. [En línea]. Disponible doi: 10.1109/ITCOSP.2017.8303122.
- [54] Image Analyst, “Split image into blocks”, MATLAB wiki, 2013. Disponible en: https://matlab.fandom.com/wiki/Split_image_into_blocks
- [55] Image Analyst, “How to Crop Image into Overlapping Patches”, MATLAB Answers, 2020. Disponible en: <https://la.mathworks.com/matlabcentral/answers/523498-how-to-crop-image-into-overlapping-patches>
- [56] Image Analyst, “Image operations, skewness and kurtosis”, MATLAB Answers, 2011. Disponible en: <https://la.mathworks.com/matlabcentral/answers/15307-image-operations-skewness-and-kurtosis>
- [57] A. Uppuluri, “GLCM Texture Features”, MATLAB Central File Exchange, 2008. Disponible en: <https://la.mathworks.com/matlabcentral/fileexchange/22187-glcm-texture-features>
- [58] MathWorks, “Segmentación basada en colores utilizando la agrupación K-medias”, MATLAB Documentation. Disponible en: <https://la.mathworks.com/help/images/color-based-segmentation-using-k-means-clustering.html>
- [59] M. Veta, “Staining Unmixing and Normalization for Histopathology Images”, Image Sciences Institute, University Medical Center, Utrecht, The Netherlands, 2013. Disponible en: <https://github.com/mitkovetta/staining-normalization>

9. Anexos

Los datos extras como códigos y conjunto de imágenes se pueden encontrar en el siguiente [enlace](#). Cabe destacar, que algunos códigos fueron elaborados con ayuda de la comunidad y documentación de MATLAB, siendo estos los principales colaboradores:

- Image Analyst, siendo uno de las principales ayudantes que nos permitió los códigos para hacer Image Patching y la extracción de características estadísticas de color. [54], [55] y [56]
- Avinash Uppuluri, la persona la cual hizo el código para la extracción de características de textura de la matriz GLCM. [57]
- Con la documentación de MATLAB, nos permitió elaborar los códigos para hacer la segmentación basada en colores con el método k-means clustering. [58]
- Mitko Veta, fue la persona que colaboró con el código para hacer la normalización de la tinción [59]