



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
FACULTAD DE BIOLOGIA  
Biología de la Célula - BIO141C

**Trabajo Bibliográfico - Entrega 2:  
Aplicación del machine learning a la predicción del  
Cáncer**

by

*Nicolás Buzeta*

*José Cáceres*

*Benjamín Barrios*

*Carlos Álvarez*

*Profesora : Alicia Nogueras*  
*Tutor : Andrés Carreño*

18 de octubre de 2020

# Índice general

<b>1</b>	<b>Introducción</b>	<b>2</b>
<b>2</b>	<b>Desarrollo de la Investigación</b>	<b>3</b>
2.1	Análisis Biológico . . . . .	3
2.1.1	«Nuclear Structure, Organization, and Oncogenesis» (Rynearson y Sussman) . . . . .	3
2.1.2	«Cancer Nucleus» (Dey) . . . . .	6
2.1.3	«Nuclear Structure in Cancer Cells» (Zink y col.) . . . . .	9
2.1.4	«The Structure and Origin of Giant Nuclei in Human Cancer Cells» (Therman y col.) . . . . .	11
2.1.5	«Inside the Cell» (Madrazo y col.) . . . . .	13
2.2	Análisis Aplicación de Machine Learning . . . . .	14
2.2.1	«Machine Learning Applications in Cancer Prognosis and Prediction» (Kourou y col.) . . . . .	14
2.2.2	«Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images» (Sirinukunwattana y col.) . . . . .	17
2.2.3	«Early Lung Cancer Detection Using Nucleus Segmentation Based Features» (Kancherla y Mukkamala) . . . . .	19
2.2.4	«Prediction of Recurrence in Early Stage Non-Small Cell Lung Cancer Using Computer Extracted Nuclear Features from Digital H&E Images» (Wang y col.) . . . . .	20

# 1 Introducción

Sabemos que dentro de las células que existen en nuestro organismo hay un núcleo (células procariontes), y este núcleo tiene en su interior una organización espacial característica, la cual a su vez depende de la matriz nuclear como de la lámina nuclear. La matriz nuclear es una red de fibras que se encuentra distribuida en todo el interior de los núcleos celulares y, además de dar rigidez al núcleo, mantiene el orden en su interior. Por otro lado, la lámina nuclear es una red que al igual que la matriz, da rigidez al núcleo, pero también regula eventos como la replicación de ADN y la división celular (Alberts y col., 2010; Zink y col., 2004, p. 200). En células cancerígenas, ambas de estas estructuras se pueden ver afectadas. Esto comúnmente se refleja en un cambio morfológico del núcleo, que va de un aspecto elipsoidal a uno irregular.

La estructura del núcleo final en estas células malignas posee diferencias características en la arquitectura nuclear con una célula sana o incluso entre esas mismas células cancerígenas hay diferencias especiales. (Rynearson & Sussman, 2011). Esto conlleva la posibilidad de detectar células cancerígenas por su forma, y más aún, identificar el tipo de cancer en específico. Así, poder detectar estos cambios no solo ayudaría a encontrar el cáncer, sino también a detectar el tipo particular, aun cuando todavía no se ha formado el tumor. Un ejemplo de esto sería la detección de células cancerígenas en la orina («Urine Cytology - Mayo Clinic», s.f.); Como estas células pueden venir de un tumor no visible, se podría detectar y tratar el cáncer sin necesidad de tomar una biopsia directa.

Considerando los aspectos expuestos, si bien se pueden detectar las células cancerígenas, es necesario ir mas allá de los métodos tradicionales, ya que es necesario captar cambios minúsculos en la estructura nuclear. En base a esto, machine learning es una buena opción para ser aplicada en esta área, ya que es capaz de aprender relaciones de manera autónoma. Esto significa que es posible darle información sobre millones de células y el programa aprenderá a identificar las células malignas, todo esto sin intervención humana (más allá de la programación propia del programa). Esta tecnología posee un gran uso actualmente, y sigue en ascenso con diversas investigaciones y progresos en el funcionamiento. (Carleton y col., 2018; Kourou y col., 2015).

También es posible automatizar, hasta cierto punto, el aprendizaje y uso del machine learning. En este caso, se podrían aplicar técnicas para clasificar los núcleos dentro de las células de manera automática. De esta forma solo sería necesario extraer la célula y podríamos automáticamente encontrar el núcleo en su interior (Sirinukunwattana y col., 2016). Así, el programa solo requeriría otro método para extraer la célula, luego todo el resto del proceso se realizaría de forma automática.

## 2 Desarrollo de la Investigación

### 2.1 Análisis Biológico

#### 2.1.1 «Nuclear Structure, Organization, and Oncogenesis» (Rynearson y Sussman)

En los últimos años, el núcleo celular ha sido protagonista de las últimas investigaciones sobre los cambios que genera el cáncer en la célula. Pero, para poder explicar porque el núcleo es tan importante en la detección de cáncer, primero tenemos que hablar de sus componentes, estructura y funciones.

En general, el núcleo está encargado de mantener o «cuidar» al genoma y regular la expresión de genes. Específicamente hablando, el núcleo se centra en realizar la transcripción y de transportar el RNA hacia el citoplasma para que pueda ser traducido en proteínas que cumplan funciones específicas que la célula necesita. El cáncer afecta estos procesos, alterando la ubicación de los cromosomas. Esto provoca que cambie la interacción entre genes y que al mismo tiempo los procesos de transcripción y traducción tengan cambios drásticos. Lo que afecta la funcionalidad de las proteínas que se crean al final de este proceso. Con esto en mente, podemos estudiar el proceso que hay detrás de la regulación de los genes y encontrar los componentes que son afectados específicamente por el cáncer, lo que nos ayudará en gran medida para desarrollar nuestro algoritmo de machine learning.

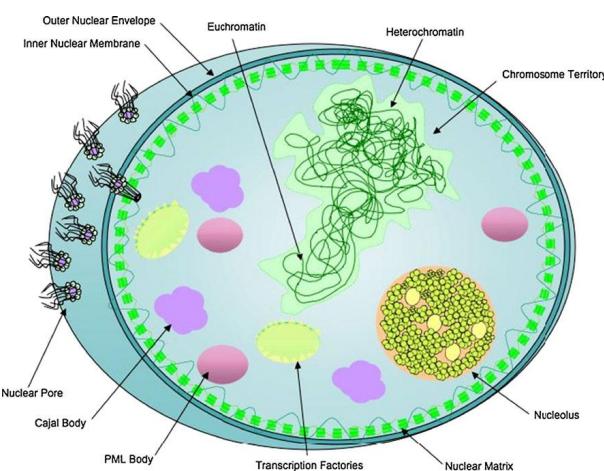


Figura 1: Estructura y organización nuclear  
(Rynearson & Sussman, 2011)

### **Estructura nuclear:**

En el núcleo hay 3 estructuras importantes: la envoltura nuclear, la matriz nuclear y los poros nucleares. La envoltura nuclear, al igual que la membrana plasmática de la célula, está compuesta de una bicapa lipídica de fosfolípidos, la cual separa el citoplasma del contenido del núcleo. Lo que lo diferencia de la membrana plasmática son las proteínas transmembrana y los poros nucleares, que tienen como función seleccionar que entra y que sale del núcleo. Otras funciones de la envoltura nuclear están relacionadas con la organización de la cromatina y la regulación de la expresión genética. También hay una diferencia entre las proteínas transmembrana de la capa externa e interna de la envoltura nuclear, las primeras interactúan con el retículo endoplasmático rugoso (se encuentra cerca de la envoltura nuclear externa) mientras que las segundas solo interactúan directamente con la matriz nuclear.

Por otro lado, la matriz nuclear está compuesta por láminas y proteínas que forman un citoesqueleto dentro de la envoltura nuclear. Por lo tanto, la matriz nuclear existe para proporcionar soporte y estructura al núcleo, dándole un lugar de anclaje al ADN, poros nucleares y proteínas nucleares. Dentro de la matriz nuclear, podemos encontrar una zona llamada espacio intranuclear donde encontramos al genoma. El genoma está organizado en territorios cromosómicos donde encontramos zonas de alta y baja actividad transcripcional, específicamente la eucromatina (ubicada en el centro) y la heterocromatina (ubicada en la periferia).

## Territorios y estructura cromosómica:

Durante la mitosis, el genoma condensado se organiza en cromosomas, los cuales se separan en la división mitótica. Después de la mitosis los cromosomas se desenrollan, formando los territorios cromosómicos. A los territorios cromosómicos que están muy cerca entre ellos, se les llama en quote vecindario cromosómico y dentro de este «vecindario» los territorios cromosómicos interactúan generando una transcripción o translocación genética coordinada. Dentro de los cromosomas también encontramos una distribución (no estricta), donde las bases CG se ubican centralmente mientras que las bases AT se ubican en la periferia.

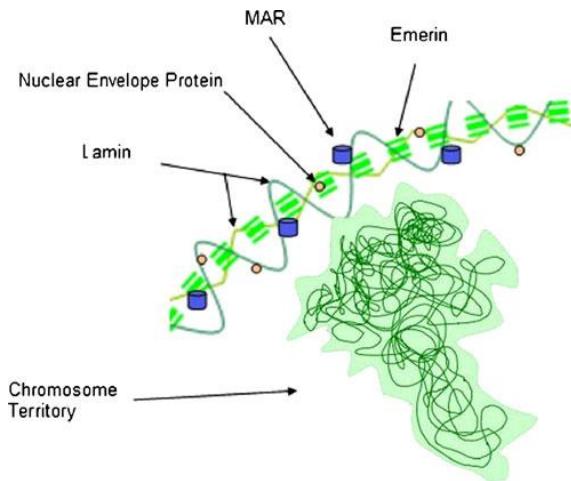


Figura 2: Matriz nuclear, incluyendo los lamine y proteínas asociadas  
(Rynearson & Sussman, 2011)

## 2.1.2 «Cancer Nucleus» (Dey)

Hay una gran cantidad de alteraciones morfológicas significativas en el núcleo de las células cancerígenas que pueden ser detectadas por microscopía de luz y con método de tinción. Estas alteraciones suelen estar asociadas a funciones celulares trastornadas de la célula cancerosa. Es por esto que la morfología de los núcleos tiene gran importancia en los estudios, ya que permite comprender los cambios que se producen en el núcleo durante el proceso de carcinogénesis de una manera muy anticipada para así interiorizar en la organización del núcleo del cáncer y ayudar significativamente a entender la patología del núcleo del cáncer y a desarrollar un diagnóstico.

El núcleo de la célula cancerosa muestra muchas características con cambios microscópicos como la alteración de tamaño nuclear, forma, margen, patrón de cromatina, nucleolos, y el espacio perinuclear. De acuerdo con algunos autores, «Los cambios progresivos de la estructura nuclear en la malignidad es probable que se asocie con cambios de la matriz nuclear (NM), anomalías de la membrana nuclear, y la organización de la cromatina nuclear.» (Nickerson, 1998)

El tamaño en núcleos malignos suele mostrar una variación que es conocida como pleomorfismo. La base de esta variación de tamaño según la teoría del fenotipo de los mutantes se explica en que, si hay inestabilidad genética, esta se produce en el inicio de la carcinogénesis, acá lo que sucede es que las células eligen la mutación óptima para la progresión del cáncer. Por lo tanto, se generan varios subclones de células con esta mutación específica. Es aquí donde estos subclones igual pueden tener relación con los cambios morfológicos variables de las células iniciales, lo que genera una población de células similares con cambios de comportamiento, forma y tamaño que las caracterizan.

Al igual que la forma del núcleo, el margen nuclear cambia en células cancerígenas, estas alteraciones pueden ser vistas como un ranurado nuclear o con convulsiones nucleares. También se ve alterado el engrosamiento del margen el cual aumenta su tamaño a menudo en los núcleos estudiados y esto se debe a la formación de heterocromatina periférica, la cual es una forma de los genes empaquetados en la cromatina de manera más compacta e inactiva. Estas se ubican en la periferia nuclear de las células cancerígenas, por lo que es un buen indicador de un cáncer temprano.

Si bien el núcleo en sí se ve afectado en células cancerígenas también organelos dentro de él se ven alterados y pueden ser de gran ayuda como indicador de un posible cáncer inicial. Uno de estos organelos es el nucleolo el cual tiene como función la biogénesis del ribosoma, esta posee una tasa que depende del metabolismo de la célula. Aquellas células que proliferan adquieren un aumento en la demanda de proteínas, lo que se traduce en un aumento de la tasa de biogénesis ribosomal. De modo que las células que más proliferan tendrán un nucleolo de mayor tamaño que deberá suplir la alta tasa de

la biogénesis del ribosoma. Se entiende igual que una célula cancerosa posee una gran tasa de crecimiento por lo tanto se espera que estas tengan nucleolos con un tamaño mayor al normal. Aunque por otro lado según Dey:

El agrandamiento nucleolar es por parte de la célula cancerígena, sin embargo, no puede explicarse únicamente por el aumento de la proliferación celular. La neoplasia intraepitelial prostática es un tumor de crecimiento lento, pero muestra grandes y prominentes nucleolos, mientras que la neoplasia intraepitelial cervical es un tumor de rápido crecimiento con nucleolos poco visibles. Esto indica que la tasa de proliferación celular puede no ser el único factor del cambio nucleolar. (Dey, 2010, p. 384)

La matriz nuclear es otro de los distintos indicadores para la detección temprana del cáncer dentro del núcleo o que relaciona al núcleo. Esta matriz nuclear es la estructura de soporte tridimensional del núcleo de la célula que realiza funciones, como determinar la morfología nuclear, la organización del ADN a lo largo del ciclo celular y la estabilización de ADN durante la replicación. Esto le otorga un gran efecto en la morfología del núcleo y como se ha descrito en párrafos anteriores el pleomorfismo está muy ligado a células cancerosas. Por lo tanto, conocer y tener en estudio aquella estructura que se le atribuye la determinación de la morfología nuclear nos da una gran ventaja sobre la detección de células cancerígenas.

También la matriz nuclear contiene una serie de proteínas que llevan el nombre del lugar donde se encuentran (NMP) las cuales según el Pranab son «... específicas en tejidos y enfermedades y por lo tanto son útiles en la detección del cáncer. La detección del NMP22 en la orina en el carcinoma de células transicionales de la vejiga es muy útil en la proyección de los casos recurrentes.» (Dey, 2010, p. 386)

Son estos indicadores los cuales no son todos los que existen, los que a través de su influencia en la morfología del núcleo entregan una significativa ayuda en la detección de células cancerosas en un periodo de tiempo de proliferación corto.

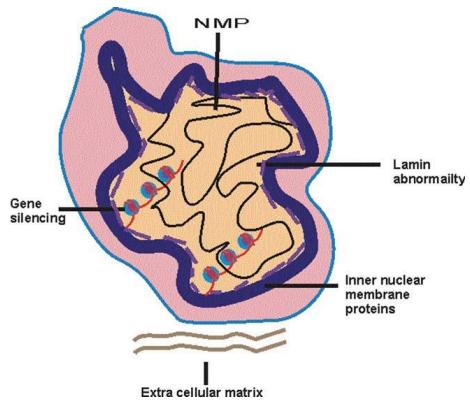


Figura 3: Diagrama esquemático muestra los diferentes factores posibles responsables por las irregularidades del margen nuclear

(Dey, 2010)

### **2.1.3 «Nuclear Structure in Cancer Cells» (Zink y col.)**

Entre los cambios morfológicos ya mencionados no tanto en el núcleo en sí, si no que dentro de él, está el que se relaciona con los cromosomas, sus translocaciones y su textura. Estas alteraciones pueden ser tan características de un tipo de cáncer dado y de la etapa en que se utilizan en el diagnóstico del cáncer.

La detección de las translocaciones cromosómicas es un aspecto importante del diagnóstico del cáncer, especialmente en lo que respecta a los cánceres hematológicos, ya que las translocaciones específicas suelen ser características de leucemias o linfomas. Las translocaciones específicas también pueden determinar el pronóstico del paciente. Por ejemplo, Zink explica que:

la translocación t(14;18) (q32;q21), que afecta al gen BCL2, se observa en alrededor del 90% de los linfomas foliculares, que por lo general siguen un curso indolente. Por el contrario, la translocación t(11;14) (q13;q32) es una característica del linfoma de células del manto, que es una enfermedad de rápida progresión. (Zink y col., 2004, p. 681)

También es importante diagnosticar los cambios en la textura de la cromatina, que se observan con frecuencia en las células cancerosas. Estos son probablemente causados por la condensación y descondensación de los dominios de cada cromatina. De acuerdo con algunos autores, «las diferencias en el grado de condensación de la cromatina y en la organización de la cromatina de las células normales y tumorales también están indicadas por las diferentes distancias entre los loci de los genes. » (Lukásová y col., 2004)

Dos configuraciones de cromatina fácilmente observables con importancia diagnóstica son el «engrosamiento de la cromatina» y la «cromatina abierta» exagerada, que corresponden a un aumento o a una pérdida de agregados de heterocromatina, respectivamente. El engrosamiento de la cromatina es poco común en las células normales y da lugar a mayores agregados de heterocromatina. Puede ser inducido en diferentes tipos de células por la expresión de HRAS activada. El grado de engrosamiento de la cromatina no está relacionado con la fracción de células que proliferan en las líneas celulares, pero se correlaciona estrechamente con el potencial metastásico. De este modo se puede estudiar el engrosamiento de la cromatina como motivo para una posible metástasis.

Table 1   Nuclear shape changes associated with cancer	
Nuclear characteristic	Cancer type
<b>Changes in nuclear shape</b>	
Malformed nuclei with frequent crush artifacts	Small-cell lung carcinoma
Grooves and long clefts	Papillary thyroid carcinoma*; Langerhans-cell histiocytosis; urothelial tumours, including Brenner tumour of the ovary; granulosa-cell tumour of ovary; follicular lymphomas
Polylobulation (nucleus subdivided into several lobes)	Some adenocarcinomas
Indentations, undulations, folds	Many kinds of cancer
<b>Chromatin changes</b>	
Coarse aggregates of heterochromatin	Many forms of cancer and precancerous changes, including follicular thyroid adenoma*
Asymmetric aggregates of heterochromatin	Wide range of cancers
Dispersed heterochromatin	Small-cell lung carcinoma
Loss of heterochromatin aggregates	Papillary thyroid carcinoma*; wide range of other carcinomas
<b>Nucleolar alterations</b>	
Enlarged nucleoli	Wide range of cancers
Inconspicuous nucleoli	Precancerous squamous lesions of the cervix; small-cell anaplastic lung carcinoma
Marked cell to cell variation in numbers or sizes	Wide range of cancers

Diagnostic changes in cancer-cell nuclei detected by conventional cytology stains\*. Changes that have been linked to the expression of particular oncogenes.

Figura 4: Cambios en nucleos afectados  
(Zink y col., 2004)

## 2.1.4 «The Structure and Origin of Giant Nuclei in Human Cancer Cells» (Therman y col.)

El ciclo nuclear normalmente posee una alternancia entre replicación cromosómica y mitosis, que, en caso de alterarse, puede producir diversas modificaciones en la mitosis y por ende anomalías cromosómicas características de cáncer. Esto puede traer poliploidía en los núcleos debido a diversas razones ligadas a la mitosis, como la supresión del huso. Esto se ha estudiado bastante en invertebrados, plantas y protistas, pero en mamíferos la información estudiada solo se reduce a modificaciones mitóticas en tumores malignos. En la mayoría de los cánceres existen los llamados núcleos gigantes, pero es importante señalar, que las llamadas modificaciones mitóticas comunes en el cáncer también se han encontrado en células normales.

En este estudio se eligieron dos carcinomas invasivos de células en el cuello uterino, ya que eran prominentes para un estudio detallado. En los tumores los nucleos iban de pequeños a grandes, donde los grandes no se pudieron dividir. Los nucleos variaron de estar teñidos de manera uniforme a tener cromocentros prominentes. Ninguna celula de los tumores mostró cromatina X identificable (esta cromatina es naturalmente imposible de determinar en núcleos con cromatina prominente). Los dos tumores tenían núcleos gigantes, pero eran diferentes en varios aspectos. En el caso 1 casi todos los núcleos presentaban cromocentros prominentes donde parece ocurrir principalmente la endomitosis (por ende poliploidía).

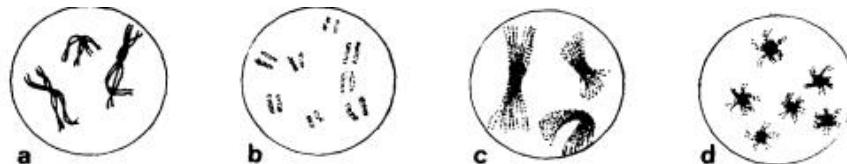


Figura 5: Diagrama de diferentes tipos de endopolyploid nuclei: (a) profase con diplocromosomas, (b) endometafase-endoanafase, (c) núcleo con cromosomas politénicos, (d) núcleo con cromocentros grandes

(Therman y col., 1983)

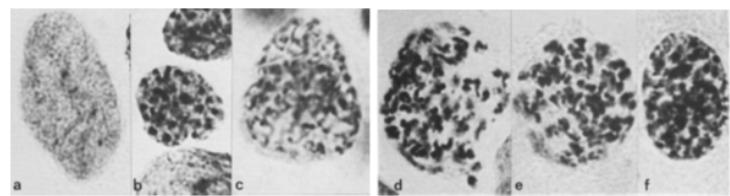


Figura 6: Diagrama de núcleos cancerosos: (a) núcleo sin cromocentros o cromatina X, (b) núcleo con cromocentros, (c) profase normal, (d) probable endoprofase, (e, f) endometafase-endoanafase (caso 1). x 3000

(Therman y col., 1983)

## 2.1.5 «Inside the Cell» (Madrazo y col.)

El núcleo separa y contiene el genoma del citoplasma, las células cancerígenas presentan una morfología nuclear aberrante, ya sea en forma y volumen, con regiones de cromatina aberrantes. A pesar de eso estos defectos en los componentes nucleares, no son conductores al desarrollo del cáncer, pero sí una consecuencia de la evolución de este. Las células cancerosas manifiestan muchas anomalías epigenéticas, que conducen a inestabilidad genómica y expresión génica aberrante durante la progresión y la recurrencia del cáncer. Debido a la arquitectura nuclear aberrante de las células cancerosas, los cambios nucleares se han propuesto como un sello distintivo del cáncer y pueden conducir a la identificación de nuevos desarrollos terapéuticos.

Las integrinas son proteínas receptoras transmembrana del tipo 1, poseen un dominio extracelular, una región transmembrana y una cola citoplasmática, se encuentran presentes en diversos procesos desempeñando un papel fundamental. Debido a estas mismas funciones es que se ven implicadas en diversas patologías, entre ellas el cáncer. Así, la mediación de integrinas pareciera impulsar la proliferación de células cancerígenas y metástasis. De esta forma las células cancerígenas utilizan las integrinas para regular la señalización del factor de crecimiento o en otros casos ayuda a mantener la población del cáncer estableciendo una relación entre las integrinas y el recién mencionado. Para las células cancerosas, se ha demostrado que las integrinas y sus glicoproteínas asociadas apoyan la adhesión de las células cancerosas y contribuyen a la progresión tumoral y a los fenotipos de cáncer invasivo. Es ampliamente conocido que las integrinas son críticas para la migración de células cancerosas en espacios confinados. La regulación de la forma, posición y deformación nuclear es fundamental para controlar la migración de células tanto normales como cancerosas.

La deformabilidad nuclear es fundamental para permitir la migración de células cancerosas en condiciones confinadas . Durante la migración celular, el núcleo se vuelve más maleable y sensible a las fuerzas aplicadas desde el citoesqueleto. Múltiples componentes nucleares determinan la deformabilidad nuclear, incluidas las láminas, la cromatina y el nucleoplasma.

## 2.2 Análisis Aplicación de Machine Learning

### 2.2.1 «Machine Learning Applications in Cancer Prognosis and Prediction» (Kourou y col.)

Dentro de machine learning tenemos dos fases principales, la estimación de dependencias desconocidas en un sistema, y el uso de estas estimaciones para producir outputs nuevos dado inputs desconocidos. Para realizar estos procesos, tenemos dos tipos principales de algoritmos, supervisado y no supervisado. En el proceso supervisado, uno entrega información con un output conocido. Un ejemplo de esto podría ser, las medidas de un tumor como input y el potencial de crecer como output. En el no supervisado, uno solo entrega información sin saber su output. Por lo cual el algoritmo sólo puede encontrar patrones o grupos dentro de la información dada.

Un ejemplo de clasificación lo podríamos ver en este gráfico:

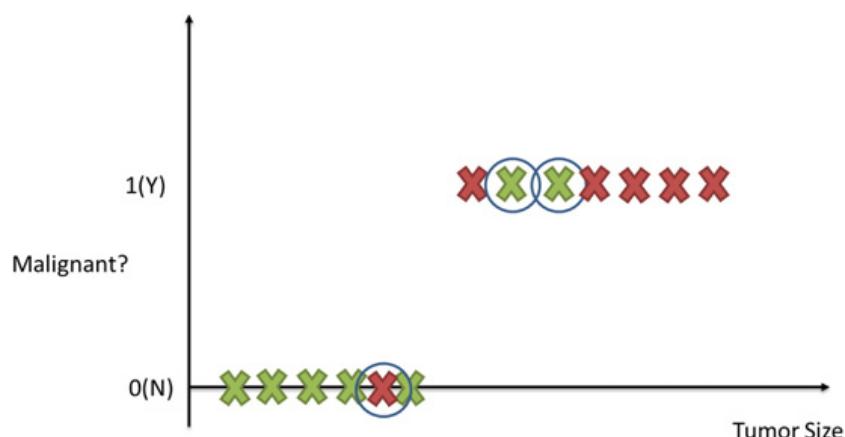


Figura 7: Grafico relacionando el tamaño de un tumor con el estado de malignidad del tumor

(Kourou y col., 2015)

Donde podríamos aplicar un algoritmo, para encontrar una línea que separe nuestros datos conocidos. Y de esa forma, cuando tengamos nuevos datos, podemos automáticamente clasificar.

Donde podríamos llegar a:

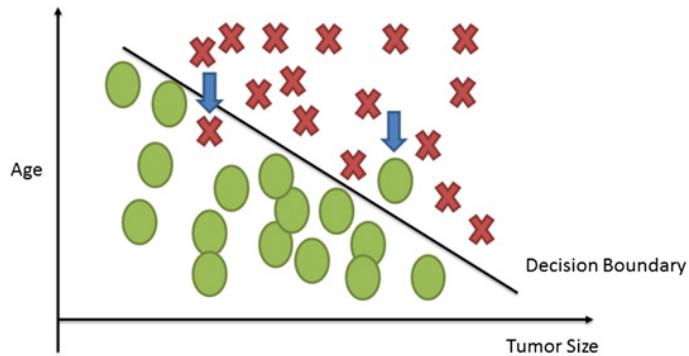


Figura 8: Grafico relacionando el tamaño de un tumor y edad del paciente con el estado de malignidad del tumor

(Kourou y col., 2015)

Donde ahora estamos trabajando con 2 fuentes de información, tamaño y edad. Por lo cual podemos ver que este método lo podemos generalizar fácilmente a diferentes dimensiones.

Otro método más reciente es denominado «ANN» y consiste de una «caja-negra» donde nosotros no sabemos como llega al resultado. Este método es muy parecido a como nosotros pensamos, ya que está modelado desde cómo funcionan las neuronas. Consiste en capas de nodos unidos por conexiones arbitrarias entre ellas.

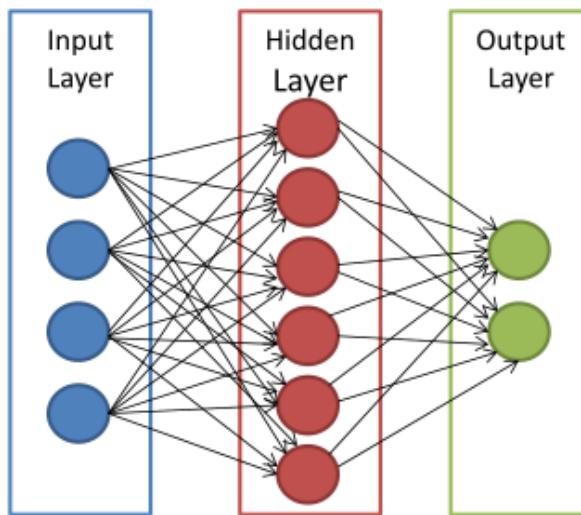


Figura 9: Grafico de un estructura general de un ANN  
(Kourou y col., 2015)

Finalmente, tenemos el algoritmo de árbol, el cual funciona en clasificar nuestra información por cada dimensión que le damos. De forma que, por ejemplo, aprende a reconocer que si tienes cierta edad y, al mismo tiempo, tienes una deformación de tu núcleo, puede predecir que tienes un cáncer peligroso.

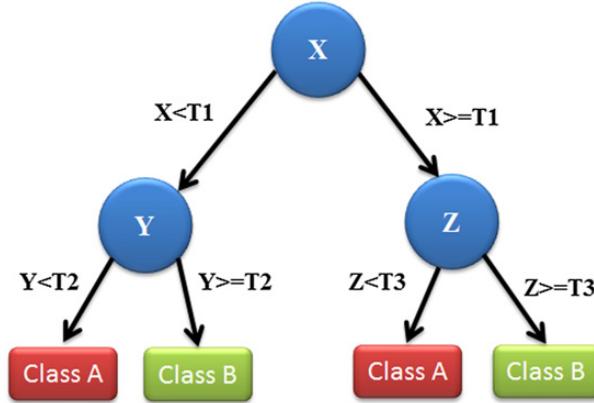


Figura 10: Grafico de un estructura general de un Arbol  
(Kourou y col., 2015)

## 2.2.2 «Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images» (Sirinukunwattana y col.)

En el uso de machine learning para cualquier tipo de aplicación, lo más importante es la data que nosotros le entregamos a nuestro modelo. Por lo que, en su uso en detección de cáncer por medio de irregularidades nucleares, vamos a tener dos problemas. La calidad de nuestros núcleos segmentados y la cantidad que podemos procesar. Este paper nos entrega un método nuevo de como usar machine learning para habilitar la detección y segmentación automática de estos núcleos.

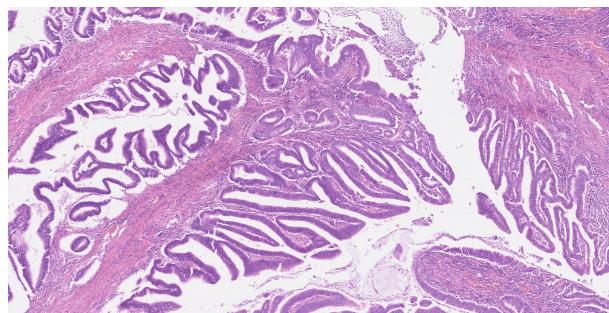
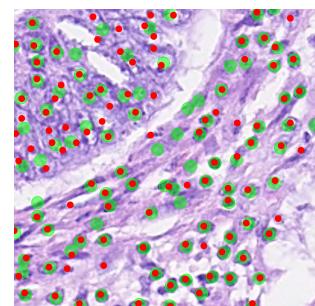
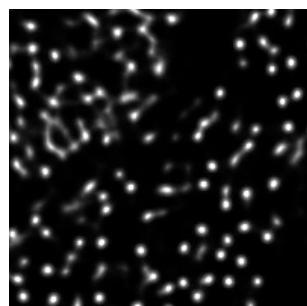
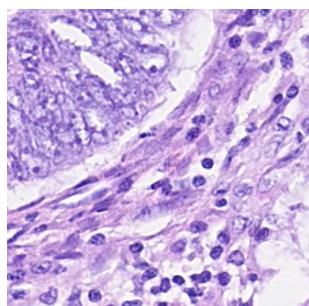


Figura 11: Foto H&E  
(Kourou y col., 2015)

Para esto usamos un modelo CNN que es un modelo derivado del ANN visto en el paper (Kourou y col., 2015). Este modelo es capaz de tomar nuestras imágenes H&E (hematoxylin and eosin), e.j. figura 11, y aprender donde están los núcleos. Esto se realizó con 100 imágenes H&E de 500 x 500 píxeles. Donde expertos y estudiantes clasificaron manualmente los núcleos presentes en cada imagen. Donde, en total, se anotaron más de 20 000 núcleos individuales.



(a) Input

(b) Mascara

(c) Núcleos marcados

Usando el método descrito anteriormente, se logró llegar a una precisión de 0.781, lo cual significa que de cada 1000 puntos clasificadas, 781 eran verdaderamente núcleos. También se logró llegar a un recall de 0.827, lo cual significa que de cada 1000 núcleos

presentes, se clasificaron 827. Esto es relevante ya que algoritmos pasados tenían precisión y recall en los alrededores de 0.600 cada uno. Por lo que las estadísticas obtenidas son una mejora.

Con este algoritmo se demuestra que es posible clasificar núcleos de células automáticamente. Lo que nos dice, que entrenar un modelo que usó características de estos núcleos no sería muy costoso. Por lo que, no habilita el uso de muchos algoritmos, los cuales, algunos, serán mencionados más abajo.

### **2.2.3 «Early Lung Cancer Detection Using Nucleus Segmentation Based Features» (Kancherla y Mukkamala)**

Actualmente uno de los pilares más importantes a la hora de combatir al cáncer, es la detección temprana y análisis correcto del tipo de cáncer, detección de metástasis, características del microambiente tumoral, etc. Para lograr esto de la forma más eficiente y rápida, se han usado un sin fin de nuevas tecnologías, pero la que ha dado mejores resultados hasta ahora, es el uso de machine learning.

El cáncer de pulmón es el cáncer que genera más muertes («Common Cancer Types - National Cancer Institute», 04/21/2015 - 08:00), por lo que ha atraído mucho atención. También se debe mencionar que se busca la detección temprana del cáncer, ya que esto sube de manera significativa la posibilidades de sobrevivencia del paciente («WHO | Early Detection of Cancer», s.f.). En este estudio se utilizaron las características del núcleo (de células tumorales) para crear un algoritmo de machine learning que logre realizar una detección temprana del cáncer de pulmón.

Para llevar a cabo esta investigación se hizo un experimento con muestras de esputo (secreciones provenientes del bronquios) de 15 pacientes diagnosticados recientemente con cáncer pulmonar y 13 pacientes normales pero que eran fumadores empedernidos (un paquete de cigarros al día). Para observar las muestra se utilizó un etiquetado de TCPP (fluorescente), que aumentó el crecimiento de lipoproteínas particulares del revestimiento de la superficie de células cancerígenas, por lo que luego se utilizó un microscopio ultravioleta (con un filtro FITC) para observar las células de las muestras de esputo. Luego se utilizaron técnicas de procesamiento de imágenes (de las muestras antes mencionadas) para separar y obtener células individuales, y de estas obtener características que sirvan en la detección de cáncer. Las características iniciales que se obtuvieron fueron: intensidad/color, forma y textura. El set inicial contenía en total 71 características pero luego de una optimización y uso de métodos capaces de identificar un set con mayor precisión, se obtuvieron 79 características. Para el uso de técnicas de machine learning se utilizó este set de características modificado (79 características) y un conjunto de datos con 60 muestras de cáncer y 59 con muestras normales (119 puntos de data en total). Los resultados finales entregaron una precisión de 87% con lo que se concluyó que los métodos de machine learning basados en el núcleo poseen un gran potencial y eficacia (comparado con los demás métodos existentes) en la detección temprana de cáncer, permitiendo mejorar el monitoreo del tratamiento, detección de recurrencia de cáncer pulmonar e identificación de pacientes que necesiten un procedimiento de diagnóstico invasivo.

## **2.2.4 «Prediction of Recurrence in Early Stage Non-Small Cell Lung Cancer Using Computer Extracted Nuclear Features from Digital H&E Images» (Wang y col.)**

Por el paper (Kancherla & Mukkamala, 2013), podemos ver que es factible usar machine learning para detectar el cáncer de manera precisa y temprana. En este estudio, el cual fue realizado en Estados Unidos, se analiza la recurrencia del cáncer de pulmón. Se utilizaron las imágenes patológicas que se obtuvieron luego de hacerles un examen radiográfico a pacientes, a los cuales ya se les había removido un cáncer de pulmón. Con estas imágenes se utilizó un algoritmo de deep learning capaz de identificar y detectar las características principales que tiene un cáncer de recurrencia.

Este algoritmo utilizó 15 «features», lo cual significa que se le entregó 15 puntos de información por paciente. Estos puntos de información se pueden separar en 3 categorías principales:

- Forma del núcleo
- Orientación del núcleo
- Textura del núcleo

Con esto, el algoritmo se encargó de poder clasificar a los pacientes en dos grupos principales. Los de no recurrencia y los de recurrencia. Para esto, se tomó el grupo colectivo de paciente y se separaron en 3 grupos. El primer grupo se usó para entrenar el algoritmo y los otros dos se usaron para revisar «accuracy» del algoritmo.

En conclusión, el estudio pudo llegar a una exactitud de 82% lo cual significa, que de cada 100 imágenes, se clasificaron 82 de manera correcta. Dado que estas imágenes se deben tomar independientemente de si se utilizan para algún tipo de algoritmo, ya que se necesitan para que los médicos revisen visualmente al paciente , se está efectivamente aprovechando el trabajo ya hecho. Por lo que predecir con el algoritmo, que tiene un 82% de exactitud, quién va a tener una recurrencia va a ser de mucha ayuda para los médicos y pacientes.

## Referencias

- Alberts, B., Wilson, J. & Hunt, T. (2010). *Biología molecular de la célula*. Barcelona, Omega  
ga  
OCLC: 708275267.
- Carleton, N. M., Lee, G., Madabhushi, A. & Veltri, R. W. (2018). Advances in the Computational and Molecular Understanding of the Prostate Cancer Cell Nucleus. *Journal of cellular biochemistry*, 119(9)pmid 29923622, 7127-7142. <https://doi.org/10.1002/jcb.27156>
- Common Cancer Types - National Cancer Institute. (04/21/2015 - 08:00). Consultado el 18 de octubre de 2020, desde <https://www.cancer.gov/types/common-cancers>
- Dey, P. (2010). Cancer nucleus: Morphology and beyond. *Diagnostic Cytopathology*, 38(5), 382-390. <https://doi.org/10.1002/dc.21234>  
\_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/dc.21234>
- Kancherla, K. & Mukkamala, S. (2013, abril). Early Lung Cancer Detection Using Nucleus Segmentation Based Features, En *2013 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. 2013 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). <https://doi.org/10.1109/CIBCB.2013.6595393>
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8-17. <https://doi.org/10.1016/j.csbj.2014.11.005>
- Lukášová, E., Kozubek, S., Falk, M., Kozubek, M., Zaloudík, J., Vagunda, V. & Pavlovský, Z. (2004). Topography of genetic loci in the nuclei of cells of colorectal carcinoma and adjacent tissue of colonic epithelium. *Chromosoma*, 112(5)pmid 14722711, 221-230. <https://doi.org/10.1007/s00412-003-0263-3>
- Madrazo, E., Conde, A. C. & Redondo-Muñoz, J. (2017). Inside the Cell: Integrins as New Governors of Nuclear Alterations? *Cancers*, 9(7)pmid 28684679. <https://doi.org/10.3390/cancers9070082>
- Nickerson, J. A. (1998). Nuclear dreams: The malignant alteration of nuclear architecture. *Journal of Cellular Biochemistry*, 70(2), 172-180. [https://doi.org/10.1002/\(SICI\)1097-4644\(19980801\)70:2<172::AID-JCB3>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-4644(19980801)70:2<172::AID-JCB3>3.0.CO;2-L)
- Rynearson, A. L. & Sussman, C. R. (2011). Nuclear structure, organization, and oncogenesis. *Journal of Gastrointestinal Cancer*, 42(2)pmid 21286858, 112-117. <https://doi.org/10.1007/s12029-011-9253-5>
- Sirinukunwattana, K., Raza, S. E. A., Tsang, Y.-W., Snead, D. R. J., Cree, I. A. & Rajpoot, N. M. (2016). Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. *IEEE Transactions on Medical Imaging*, 35(5), 1196-1206. <https://doi.org/10.1109/TMI.2016.2525803>

- Therman, E., Sarto, G. E. & Buchler, D. A. (1983). The structure and origin of giant nuclei in human cancer cells. *Cancer Genetics and Cytogenetics*, 9(1), 9-18. [https://doi.org/10.1016/0165-4608\(83\)90019-5](https://doi.org/10.1016/0165-4608(83)90019-5)
- Urine Cytology - Mayo Clinic.* (s.f.). Consultado el 6 de septiembre de 2020, desde <https://www.mayoclinic.org/tests-procedures/urine-cytology/about/pac-20385279>
- Wang, X., Janowczyk, A., Zhou, Y., Thawani, R., Fu, P., Schalper, K., Velcheti, V. & Madabhushi, A. (2017). Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H&E images. *Scientific Reports*, 7(1), 13543. <https://doi.org/10.1038/s41598-017-13773-7>
- WHO | Early Detection of Cancer.* (s.f.). Consultado el 18 de octubre de 2020, desde <http://www.who.int/cancer/detection/en/>
- Zink, D., Fischer, A. H. & Nickerson, J. A. (2004). Nuclear structure in cancer cells. *Nature Reviews Cancer*, 4(9), 677-687. <https://doi.org/10.1038/nrc1430>