



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUOLA TÉCNICA
SUPERIOR INGENIERÍA
INDUSTRIAL VALENCIA

Curso Académico:

Agradecimientos

Desde que tengo uso de razón, mis padres me han enseñado a agradecer a los demás los esfuerzos y trabajos que hacen por mí, ya que somos el resultado de todo lo que han hecho por nosotros. Este trabajo, que ha resultado ser el primer trabajo serio que llevo a cabo, además del primer reto en el ámbito académico que enfrento, no iba a ser menos.

Antes de nada, agradecer a mis tres tutores JuanMi, Miguel Ángel y Javi, por su gran dedicación en todas las etapas del proyecto. Los tres han sido fundamentales para mí para poder desarrollar este estudio, además de haberme otorgado durante el proceso todas las herramientas necesarias para ser un buen investigador. JuanMi por sus dotes de organización y gestión de proyectos, Miguel Ángel por sus dotes dentro de los campos de las ciencias exactas como la matemática, estadística y física y Javi por sus dotes en el campo de la informática y la inteligencia artificial.

En primer lugar, me gustaría agradecer a mi familia por todo el apoyo que me han brindado en los momentos mas bajos y de estrés, por animarme a seguir estudiando lo que me gusta y por celebrar mis logros con el mismo entusiasmo que si fueran suyos propios. No solo hablo de mi familia biológica, sino también de mi familia putativa, mis tíos y tías que, sin compartir sangre con ellos, los considero de mi propia familia.

Agradezco especialmente a mis padres, Teresa y Alberto, los cuales me han apoyado día a día, ayudándome a organizar mis tareas, nunca presionándome por sacar mas o menos notas, celebrando tanto mis matriculas como mis aprobados raspados y dándome mensajes de tranquilidad cuando el estrés se apoderaba de mí. No solo me han ayudado en el ámbito académico, sino que, en parte, me han formado como la persona que soy hoy en día, permitiéndome viajar, salir con mis amigos sin ninguna restricción y dándome una educación excelente basada en el respeto, el trabajo duro y la generosidad.

Me gustaría también dentro de mi familia agradecer a mi prima Lola su inestimable ayuda desinteresada en el proceso de maquetación y presentación de este proyecto, lo que me ha ahorrado muchas noches en vela y posibles episodios de estrés.

En segundo lugar, me gustaría agradecer a mis amigos de toda la vida, consiguiendo motivarme día tras día para seguir estudiando esta carrera, convirtiendo mis dramas en tardes y noches de risas y diversión.

En tercer lugar, me gustaría darle las gracias a mis amigos de la carrera, los cuales han conseguido que cada día vaya con mas ganas a clase para verlos y poder disfrutar con ellos, además de haberse convertido en un apoyo incondicional para mí, consiguiendo de ellos amistades tan fuertes como si los conociera desde el primer día que pude hacer amigos.

En cuarto lugar, me gustaría agradecer a mis compañeros de karate, los cuales me han ayudado a crecer como persona en esta disciplina, ya que sin ellos jamás podría haber podido progresar en el

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

karate, una disciplina que mucho mas que ejercitarme me ha otorgado una filosofía y actitud frente a la vida que ha sido crucial en mis estudios.

En quinto lugar, me gustaría agradecer a Lena y Miguel Angel por todo lo que me han enseñado, por todas las tardes hablando y planeando las estrategias del proyecto, por todas las enseñanzas que me han dado, tanto en el campo matemático como físico, por el interés que han puesto en mi y sobre todo por lo acogido que me han hecho sentir, formando los 3 un gran y familiar equipo de trabajo. Muchas gracias y merci beaucoup.

Por último, me gustaría agradecer especialmente este trabajo a mis maestros, si, digo maestros, porque considero que la palabra profesores se les queda cortos, ya que me han enseñado mucho mas que un temario.

Luis Mata, mi maestro de matemáticas, muchas gracias por despertar mi amor hacia las matemáticas desde una temprana edad, y hacerme ver que valgo para ellas, confiando en mi día tras día. Nunca tuviste intención de enseñarnos el temario de selectividad, sino de enseñarnos matemáticas de verdad, al igual que nunca quisiste que aprobáramos o suspendiéramos exámenes, sino que entendiéramos las matemáticas y todas sus aplicación en los diferentes escenarios de la vida. Además de motivar a tus alumnos día tras día y hacernos sentir como si fuéramos tus sobrinos. Siempre me pregunte como podía haber un corazón tan grande en tu pecho.

Maite Beltran, mi maestra de biología, mai podria deixar d'agrair-te tot el que vas fer per mi, vas ser la causant de què m'enamorara de les ciències de la vida i de què tirara aquesta carrera. Sempre em motivaves a seguir estudiant biologia, en les teues curiositats, concursos, excursions, etc. No crec que ningú tinga el dret a dir que ets una professora de biologia, per a mi seria com un insult, jo diria que ets una mestra de la vida, una coneixedora de la natura o una enamorada de la seua disciplina. Si la vida acadèmica d'una persona tinguera vida pròpia, tu series la mare de la meua.

Paula, Julio, Jero y Guillermo, mis senseis de karate. Ellos me enseñaron que el karate no es un deporte, sino un arte marcial y una filosofía aplicable a cualquier ámbito de la vida. Paula, por ser la persona que consiguió engancharme a este arte y guiarme en mis primeros años, sembró en mí el interés por esta. Jero, por enseñarme el valor del esfuerzo y la repetición para alcanzar la perfección. Julio, por introducirme en el verdadero significado del karate, tratarme como si fuera un alumno mas de su clase avanzada desde el principio y por ayudarme además con todo tipo de dudas informáticas. Guillermo, sensei, por ser aquel que me ha abierto los ojos totalmente a el karate, que me ha enseñado a nunca rendirme, a ser un luchador, a creer en mi y a buscar siempre nuevos retos y horizontes.

Javier Juan, mi maestro de Informática, muchas gracias por ser la persona que me ha hecho entender de verdad como abordar un problema informático, como trabajar correctamente y como tratar los datos correctamente. Gracias a ti, aunque a final de carrera, me has enseñado como trabaja un ingeniero de verdad. Gracias por ser tan generoso conmigo, resolver mis dudas en todo momento sacando huecos para mi hasta cuando ibas mas apurado, y sobre todo muchas gracias por transmitirme todos tus conocimientos. Eres un grande tío.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Para finalizar, me gustaría dedicarme este trabajo a mi abuelo, Alberto Belenguer Lajusticia, que falleció hace pocos meses. Al igual que muchas grandes personas, sus historias se pierden, pero los que lo conocimos jamás lo olvidaremos. Fue una persona generosa, atenta con los demás, extremadamente trabajador, cariñoso, comprensivo y noble. Siempre estuvo para los demás, ayudando a todos aquellos que lo necesitaron, sin pedir nunca nada a cambio, sin ni siquiera buscar un gracias. Nunca me olvidare de cuando me llevabas al colegio y siempre querías que fuera el primero en la fila, de cuando me regalabas mis cromos de pokemon todas las semanas, de cuando me cuidabas en tu chalet todos los veranos y me llevabas a la playa, de cuando me llevabas a almorzar y pasábamos horas hablando, o más bien, tu escuchándome siempre con una sonrisa. No solo fuiste un abuelo, fuiste un padre para mis primos Lucas, Sara y María, y así es como te recordare, como abuelo, padre y amigo.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

RESUMEN

El estudio de la morfología de la membrana plasmática celular se trata de un campo que cada vez despierta más interés entre los investigadores de las diferentes ramas biológicas. Estas membranas están estructuradas en diferentes subdominios, los cuales cuentan con diferentes funciones relacionadas tanto con el metabolismo, transporte de productos y comunicación de la célula. El conocimiento de la morfología de esta membrana es por tanto un factor esencial para estudiar el comportamiento de las células. Esta problemática se encuentra estrechamente relacionada con el ámbito médico, ya que gran cantidad de enfermedades causan alteraciones en el funcionamiento y la morfología de esta capa, siendo así parte de los síntomas de estas. La detección temprana de estas alteraciones se trata de una de las bases de la llamada medicina preventiva.

Gracias al avance de las nuevas tecnologías *single particle tracking* (SPT), las cuales nos permiten hacer el seguimiento de ciertas moléculas marcadas, somos capaces de analizar la membrana celular desde un nuevo punto de vista, el del movimiento de sus componentes.

De esta forma, el objetivo de este trabajo consiste en el diseño de un sistema capaz de mapear las superficies celulares, en especial sus microdominios específicos, mediante el análisis del movimiento de sus componentes, ajustando estos a diferentes modelos estadístico-matemáticos convencionales en la literatura, donde cada uno de estos modelos está relacionado con un tipo de entorno diferente de la membrana plasmática. De esta forma, se desarrollará una herramienta útil para su posterior uso en investigación sanitaria o aplicación clínica.

Visto el interés de este campo, desde la comunidad científica especializada se ha organizado el reto Andi Challenge (*Anomalous Diffusion Challenge*) para comprar y mejorar las herramientas computacionales existentes.

Palabras clave: *Random Walks*; *Single particule tracking* (SPT); modelos difusivos; red neuronal convolucional; membrana plasmática; microdominios.

ABSTRACT

The study of the morphology of the plasmatic cell membrane is one of the main research fields that arouse interest among researchers of different biological specialties. These membranes are structured in different subdomains, which have different functions related to both metabolism, product transport and communication between cells. The knowledge of the morphology of this membrane is in fact an essential factor to study the cell behavior. This topic is closely related with the medical field, due to a lot of illnesses causing alterations in the correct functioning and the morphology of this layer, being thus part of the symptoms of these health alterations. The early detection of these alterations is one of the basis of the so called preventive medicine.

Thanks to the advances of the new SPT technologies (Single Particle Tracking) -which permit us do the tracking of certain marked molecules- we are able to analyze the cell membrane from a new point of view, the one of their components movement.

Thus, our aim will be to design a system able to map the cell membranes, specially their specific microdomains, through the analysis of their components movement, adjusting them to different statistic and mathematical models proposed by the literature, where each of these models is related to a different cell membrane environment. Therefore, we will develop an useful tool for their posterior use in research and clinical application.

Seen the interest in this field, the specialized scientific community organize the Andi Challenge (Anomalous Diffusion Challenge) to compare and improve the existing computational tools.

Key words: *Random Walks; Single particule tracking (SPT); diffusive modelss; convolutional neural networks, plasmatic membrane; microdomains.*

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

RESUM

L'estudi de la morfologia de la membrana plasmàtica cel·lular es tracta d'un camp d'estudi que cada volta desperta més interès entre els investigadors de les diferents branques biològiques. Aquestes membranes estan estructurades en diferents subdominis, els quals contenen diferents funcions relacionades tant amb el metabolisme, transport de productes i comunicació de la cèl·lula. El coneixement de la morfologia d'aquesta membrana és per tant un factor essencial per a l'estudi del comportament de les cèl·lules. Aquesta problemàtica està estretament relacionada en l'àmbit mèdic, pel fet que gran quantitat de malalties causen alteracions en el funcionament i la morfologia d'aquesta capa, sent així part dels símptomes d'aquestes. La detecció primerenca d'aquestes alteracions es tracta d'una de les bases de la nomenada medicina preventiva.

Gràcies a l'avanç de les noves tecnologies SPT (Single Particle Tracking), que ens permeten fer el seguiment de certes molècules marcades, som capaços d'analitzar la membrana cel·lular des d'un nou punt de vista, el moviment dels seus components.

D'aquesta forma, es pretén dissenyar un sistema capaç de mapejar les superfícies cel·lulars, en especial els seus microdominis específics, mitjançant l'anàlisi del moviment dels seus components, ajustant aquests a diferents models estadístic-matemàtics convencionals de la literatura, on cada un d'aquests models està relacionat amb un tipus d'entorn diferent de la membrana plasmàtica. D'aquesta forma es desenvoluparà una eina útil per al seu posterior ús en investigació sanitària o aplicació clínica.

Vist l'interès en aquest camp, la comunitat científica especialitzada ha organitzat l'Andi Challenge (*Anomalous Diffusion Challenge*) per tal de comprar i millorar les ferramentes computacionals existents.

Paraules clau: *Random Walks; Single particle tracking (SPT);* models difusius; xarxa neuronal convolucional; membrana plasmàtica; microdominis.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

DOCUMENTOS CONTENIDOS:

- MEMORIA
- PRESUPUESTO

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

MEMORIA

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS
BROWNIANOS ANÓMALOS BASADO EN REDES
CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y
SEGMENTACIÓN EN ENTORNOS CELULARES

AUTOR: ALBERT BELENGUER I LLORENS

TUTOR: JUAN MIGUEL GARCÍA GÓMEZ

COTUTORES: MIGUEL ÁNGEL GARCÍA-MARCH Y JAVIER JUAN ALBARRACÍN

CURSO: 2019-2020

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

ÍNDICE DE LA MEMÓRIA

1.	Introducción	1
1.1	Estado del arte.....	1
1.2	Motivación.....	3
1.3	Objetivos	4
2.	Marco teórico	5
2.1	Membrana plasmática.....	5
2.1.1	Estructura	5
2.1.1.1	Estructura general	5
2.1.1.1.1	Lípidos de membrana.....	6
2.1.1.1.2	Proteínas	7
2.1.1.1.3	Glicoconjugados	8
2.1.1.1.4	Colesterol	9
2.1.1.2	Microdominios	9
2.1.1.2.1	Caveolas	10
2.1.1.2.2	Balsas Lipídicas (<i>Lipid Rafts</i>).....	10
2.1.1.3	Métodos de detección de microdominios	11
2.1.1.3.1	Resistencia a Detergentes	12
2.1.1.3.2	Observación Directa	12
2.1.1.3.3	Métodos Indirectos	13
2.1.2	Funciones.....	13
2.2	Métodos de rastreo molecular.....	14
2.2.1	Tipos de métodos existentes.....	14
2.2.2	Single particle tracking (SPT)	15
2.2.2.1	Partículas de seguimiento	15
2.2.2.2	Funcionamiento del proceso	16
2.3	Modelos difusivos.....	20
2.3.1	Random walks	20
2.3.2	Mean square displacement	21
2.3.3	Tipos de difusión	23
2.3.3.1	Difusión Normal.....	23
2.3.3.2	Difusión Confinada	24

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

2.3.3.3	Difusión dirigida.....	24
2.3.3.4	Difusión anómala.....	25
2.3.3.4.1	Continuous-time Random walks (CTRW)	26
2.3.3.4.2	Fractional Brownian Motion (FBM).....	30
2.3.3.4.3	Annealed transit time model (ATTM).....	31
2.3.3.4.4	Scaled Brownian Motion (SBM)	32
2.3.3.4.5	Levy Walks.....	33
2.3.3.4.6	Resumen.....	35
2.4	Redes Neuronales Artificiales.....	35
2.4.1	Definición redes neuronales convolucionales para clasificación	36
2.4.1.1	Arquitectura de la red	37
2.4.1.1.1	Red Convolutiva.....	37
2.4.1.1.2	Red densa	39
2.4.1.1.3	Sistema de aprendizaje	41
2.4.1.2	Diseño de la red.....	43
2.4.1.2.1	Hiperparametros	44
2.4.1.2.2	Batch Normalization.....	44
2.4.1.2.3	Sobreentrenamiento	45
3.	Materiales.....	46
3.1	Código generador de trayectorias.....	46
3.2	Servidor de cómputo.....	46
3.3	Lenguajes y tecnologías.....	47
4.	Métodos	48
4.1	Minería de datos sobre trayectorias	48
4.1.1	Generación trayectorias	48
4.1.2	Generación matrices	49
4.1.3	Definición variables	51
4.1.3.1	Orden en el tiempo	51
4.1.3.2	Distancia entre saltos	52
4.1.3.3	Tiempo máximo de espera.....	53
4.1.3.4	Desviación estándar del coeficiente de difusión aproximado	54
4.1.3.5	Normalización.....	54
4.1.4	Formación de datasets	55

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

4.2	Diseño del Modelo	57
4.2.1	Estrategia de diseño de redes	57
4.2.2	Estrategia de entrenamiento	58
4.2.3	Métricas de Evaluación	59
5.	Resultados	60
5.1	Experimento 1	60
5.2	Experimento 2	61
5.3	Experimento 3	63
5.4	Experimento 4	64
5.5	Experimento 5	66
6.	Discusión	68
7.	Conclusión	72
8.	Bibliografía.....	73

ÍNDICE DE TABLAS

Tabla 1: Resumen de las características generales de los modelos.....	35
Tabla 2: Intervalos de α por modelo	56
Tabla 3: Tabla resumen del comportamiento de cada experimento en la clasificación de cada modelo y su tasa de aprendizaje	71

ÍNDICE DE FIGURAS

Figura 1: Membrana plasmática (Lizana Nuñez, 2015).....	5
Figura 2: Tipos de lípidos de membrana divididos en dos grupos principales, fosfolípidos y glicolípidos.(Lison, 2018).....	6
Figura 3: Bicapa lipídica, se observa como la parte apolar (colas en zig-zap) quedan enfrentadas entre sí, y la parte polar (bolas) quedan en el exterior (Membrana Plasmática, 2020).....	6
Figura 4: aminoácidos del grupo apolares aromáticos, varían su cadena lateral según el aminoácidos concreto (Lison, 2018).....	8
Figura 5: Molécula de colesterol completa, cada vertice corresponde con un átomo de carbono, y las aristas los enlaces covalentes entre estos (Lison, 2018).....	9
Figura 6: Caveola con sección de la membrana invaginada señalada (CajalseyGalileos, 2013)	10
Figura 7: Corte transversal de la membrana con lipid raft (2), este presenta un mayor grosor que el resto debido a sus cadenas saturadas (Universidad de Alcalá (UAH), 2011).....	11
Figura 8: Función Gaussiana 3D	17
Figura 9: Microscopio TIRF (Fogarty, Zhang, Grigsby, Johnson, Chen, Mueller & Mansky, 2011).....	18
Figura 10: Posición real vs posición estimada (Manzo & Garcia-Parajo, 2015).	19
Figura 11: Proceso completo SPT (Manzo & Garcia-Parajo, 2015).	19
Figura 12: Grafica MSD, empleando un coeficiente de difusión de 3.5027 y unas 100 trayectorias en total.	21
Figura 13: Tipos de difusión según MSD, con vector r como coordenada espacial 2D y τ como la variable temporal	23
Figura 14: Distribución exponencial y heavy-tailed, siendo x las variables que modela la función $f(x)$ en el espacio (Glen, 2016).....	28
Figura 15: Ejemplo de CTRW de 1000 pasos	30
Figura 16: Ejemplo de FBM de 1000 pasos	31
Figura 17: Ejemplo de ATTM de 1000 pasos	32
Figura 18: Ejemplo de SBM de 1000 pasos	33

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

Figura 19: ejemplo de LW de 1000 pasos.....	35
Figura 20: Red convolucional concatenada con una red densa con salida de 9 clases (clasficiacion de numeros del 0 al 9, teniendo en este caso como input el numero 4) (Shreyak, 2020).	36
Figura 21: Kernel 3*3 con 0-padding (Gimp.org, 2016).	37
Figura 22: Convolucion del kernel con la casilla central de una matriz (Gimp.org, 2016).....	38
Figura 23: Función de activación sigmoide (izquierda) y ReLu (derecha) (Sharma, 2017).	38
Figura 24: Max pooling y mean-pooling (average pooling) con secotroz de 2x2 (Rahman, 2019).....	39
Figura 25: Capas densamente conectadas (Torres, 2018).	39
Figura 26: Perceptrón con 5 entradas (Torres, 2018).	40
Figura 27: Funciones logarítmicas graficadas.....	42
Figura 28: A la derecha learning rate demasiado alto y la izquierda learning rate demasiado bajo (educative.io, 2017).....	44
Figura 29: Proceso de batch normalization sobre distribución de Possion (Ioffe & Szegedy, 2015)....	45
Figura 30: Dropout $p=0.5$ en la capa profunda (Dubbura, 2018).....	45
Figura 31: Seis trayectorias diferentes generadas con el modelo CTRW, tienen la misma longitud pero diferente forma y distribución en el espacio	49
Figura 32: Trayectoria del modelo ATTM codificada en tiempo	52
Figura 33: Grafica Accuracy experimento 1	60
Figura 34: Grafica loss experimento 1.....	60
Figura 35: Matriz de Confusión experimento 1.....	61
Figura 36: Grafica Accuracy experimento 2	61
Figura 37: Grafica Loss Experimento 2	62
Figura 38: Matriz de Confusión Experimento 2.....	62
Figura 39: Grafica Accuracy experimento 3	63
Figura 40: Grafica Loss Experimento 3	63
Figura 41: Matriz de Confusión experimento 3.....	64
Figura 42: Grafica Accuracy experimento 4	64
Figura 43: Grafica loss experimento 4.....	65
Figura 44: Matriz de Confusión experimento 4.....	65
Figura 45: Grafica Accuracy experimento 5	66
Figura 46: Grafica Loss experimento 5	66
Figura 47: Matriz de confusión experimento 5	67

1. Introducción

El estudio de la membrana plasmática se trata de un factor crucial para el desarrollo de la medicina y la biología, ya que nos permite conocer gran parte del funcionamiento de las células, en especial mediante el estudio de sus microdominios.

La gran dificultad resulta a la similitud visual de los diferentes dominios de la membrana, que en muchos casos pueden resultar idénticos morfológicamente a la vista, pero cuentan con funciones muy distintas entre si. Por tanto, es necesario el desarrollo de nuevas formas de caracterización de esta capa que no estén basadas en su observación visual, es decir, que estén basadas en sus características intrínsecas.

Una característica propia y básica de la membrana plasmática se trata de su fluidez, es decir, sus elementos están en continuo movimiento. Esta característica tan básica y a la vez tan interesante ha llevado a muchos científicos a pensar que el estudio del movimiento de los propios componentes podría desvelar la morfología de las diferentes regiones de la membrana, ya que estos elementos se comportaran de forma distinta según en qué región se encuentren.

Esta premisa se ha llevado a cabo a día de hoy gracias a la aparición de las tecnologías *Single Particle Tracking* (SPT). Estas tecnologías nos permiten hacer seguimientos individuales de moléculas marcándolas con sustancias emisoras de radiaciones y siendo captadas mediante precisos receptores. De esta forma se pueden rastrear el movimiento de los lípidos por la membrana.

Los movimientos de estos elementos se pueden agrupar en 4 grupos principalmente: movimientos Brownianos (presentan solo un componente aleatorio), dirigidos (presentan un movimiento prevalentemente en una dirección), confinado (se mueven dentro de un espacio) y anómalos (presentan un componente no-ergódico). Los 3 primeros grupos han sido descritos fácilmente mediante sencillos análisis estadísticos, siendo el cuarto tipo, los anómalos, debido a su no-ergodicidad y gran cantidad de variaciones que presenta los que no han sido aun estudiados.

Debido a esto, nuestro estudio se centrara en la caracterización de los movimientos anómalos a lo largo de las membranas plasmáticas, entendidos como medios 2D, empleando para ello sistemas de Deep learning como son las redes neuronales convolucionales 2D. Buscaremos mediante estas herramientas diseñar un sistema capaz de ajustar los movimientos anómalos a diferentes modelos de generación de movimientos aleatorios anómalos suministrados por las bases de datos del Andi Challenge, los cuales han demostrado su manifestación y aparición en medios biológicos.

1.1 Estado del arte

El estado del arte dentro de este problema es muy limitado, ya que el desarrollo de las tecnologías SPT conjuntamente con la mejora en los métodos de machine y Deep learning son muy recientes,

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

además de que el planteamiento de este problema ha pasado desapercibido durante muchos años, debido a que los científicos no disponían del conocimiento suficiente para aplicar modelos estadísticos y matemáticos a los movimientos celulares. Por tanto, este avance empieza con el desarrollo de ramas especializadas de la ciencia de datos como la bioinformática y la informática biomédica.

En la mayoría de los estudios se centran en distinguir entre los 4 modelos básicos de movimiento browniano: normal, confinada, dirigido y anómalo. Tomando un modelo estándar para el movimiento anómalo.

Es el caso del estudio llevado a cabo por Szwabiński y su equipo (Kowalek, Loch-Olszewska, & Szwabiński, 2019) el cual emplea métodos de machine learning como *random forest* (RF) o *gradient Boosting* (GB) y métodos de Deep learning como CNN para distinguir entre estos cuatro modelos, tomando como modelo anómalo el FBM proporcionado por la API de Python.

Para los modelos de machine learning se obtienen diversos parámetros por parte de los investigadores como el coeficiente de difusión estimado, el exponente anómalo (ambos estimados a partir de la curva eMSD), asimetría, eficiencia, dimensión fractal, gaussianidad, kurtosis, radio de MSD, rectitud y atrapamiento. Estos otorgan resultados muy satisfactorios con precisiones de 96'23% y 96'47% para RF y GB respectivamente, debido a que se emplean parámetros que describen muy bien los modelos, como el caso de la rectitud el modelo dirigido y el atrapamiento el confinado.

Para el modelo de *Deep learning* CNN emplea datasets de series temporales con varios canales, es decir, varios canales de una dimensión, si es 2D la trayectoria trabaja con dos canales, uno para la x y otro para la y. Entrenando este modelo durante unas pocas épocas y una arquitectura de 6 capas convolucionales y una densa consigue un accuracy del 97'30%. La red es capaz de inferir las características de estos modelos.

Otro caso similar se trata del estudio llevado a cabo por Ralf Metzler y su equipo (Thapa, Lomholt, Krog, Cherstvy, & Metzler, 2018) el cual emplea un método de clasificación bayesiana basado en el algoritmo propuesto nested-sampling, el cual permite extraer características de conjuntos de datos y emplearla para clasificarlos y distinguir entre modelos. Para este estudio han empleado tres modelos a diferenciar, en primer lugar, la clásica difusión normal, o como ellos la han nombrado, brownian motion (BM), nuevamente el modelo FBM para representar el movimiento anómalo y el modelo de diffusing diffusivity (DD), el cual se trata de un modelo BM el cual va cambiando su coeficiente de difusión a patches, por tanto, no podría considerarse como un modelo ATTM. Consiguiendo para estos modelos un accuracy máximo del 0'782 (78'2% de acierto).

Otro caso que llama la atención se trata del estudio llevado a cabo por Giovanni Volpe y su equipo (Bo, Schmidt, Eichhorn & Volpe, 2019) el cual emplea redes neuronales recurrentes (RNN) para diferenciar entre movimiento difusivo normal y anómalo, encontrando dentro del anómalo modelos subdifusivos y superdifusivos indeterminados. Este estudio se basa en la caracterización del cuando es o no normal mediante el coeficiente anómalo de la trayectoria, el cual es inferido mediante el MSD, que es inferido a su vez por medio de las redes neuronales recurrentes, las cuales son capaces de extraer parámetros basados en la ordenación temporal cuando se disponen de poca cantidad de datos

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

(como ya hemos visto anteriormente el MSD requiere de gran cantidad de datos para aproximarlos al tMSD). Además, este estudio desarrolla un método capaz de determinar el tiempo de cambio entre modelos, es decir, proporcionando una trayectoria y determinando en que instante temporal pasa de ser un movimiento normal difusivo a un anómalo. Mediante este estudio se consiguen resultados muy beneficiosos, son capaces de inferir el tiempo de cambio entre modelos con exactitud.

Algunos estudios, aunque muchos menos que los anteriores nombrados, se centran en un nuevo reto mucho más ambicioso, la diferenciación solo entre modelos de difusión anómalos, como la que se nos plantea en el challenge en el que estamos participando, estos no emplean todos los modelos que se están estudiando, pero sí parte de ellos o similares.

Encontramos el caso del estudio realizado por Gorka Muñoz, Miguel Ángel García-March y su equipo (Muñoz-Gil, Garcia-March, Manzo, Martín-Guerrero & Lewenstein, 2019), el cual tiene como objetivo principal caracterizar diferentes modelos de movimiento anómalo, tanto su coeficiente anómalo como su modelo, empleando técnicas de RF. Para ello emplea los modelos CTRW, FBM y ATTM. Estos modelos los empleamos en el challenge, además de emplear el mismo código generador debido a que este artículo es de los mismos organizadores del challenge.

Los parámetros con los cuales realizan el RF se extraen de las trayectorias de los 3 modelos después de ser normalizados respecto a su desviación estándar. Centrándonos en el apartado de determinación de los modelos, este estudio ha conseguido unos accuracys de 89'2%, 86'6% y 50'4% para los modelos CTRW, FBM y ATTM respectivamente. El bajo accuracy del ATTM según el estudio se debe a la gran similitud entre el modelo ATTM y CTRW.

1.2 Motivación

Como se puede observar en el estado del arte, la distinción entre modelos básicos de movimiento browniano, es decir, normal, dirigido, confinado y anómalo, se trata de un campo muy estudiado y que ha generado resultados muy satisfactorios empleando diferentes técnicas de machine y Deep learning. En cambio, la distinción entre distintos modelos de movimiento anómalo sigue siendo un campo muy poco estudiado, que además de contar con muy pocos estudios al respecto, cuenta con una muy reducida variedad de modelos a experimentar.

La variedad de modelos es esencial, ya que los movimientos anómalos suponen la mayoría de los movimientos de interés de las moléculas sobre la membrana celular, tanto de material genético como de moléculas dentro de microdominios. El conocimiento del tipo del modelo de movimiento de una molécula sobre un medio nos otorgará toda esa información que deseamos y que además es crucial para el estudio de la membrana celular y el diagnóstico de patologías.

Debemos por tanto mejorar estos estudios que no trabajan con los suficientes modelos, además de contar con precisiones bastante bajas en algunos casos, como se trata del caso del último estudio comentado.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Así pues, es necesario realizar un nuevo estudio con tal de desarrollar una nueva herramienta capaz de trabajar con una buena variedad de modelos matemáticos, patentes en biología celular, y que nos aporte una precisión suficientemente buena como para predecir sin lugar a dudas el tipo de movimientos. De esta forma se podrá abordar el problema de la extracción de información de la membrana celular desde un nuevo punto de vista, no solo el bioquímico, sino el cinético.

Visto que los modelos basados en machine learning como el RF, en muchos casos son incapaces de extraer características suficientemente descriptivas de los datos, optamos por emplear nuevos sistemas más especializados y complejos basados en Deep learning. Los cuales, confiamos que mediante un correcto procesamiento de los datos y entrenamiento serán capaces de aportar mejores resultados que los que se tienen hasta ahora.

La idea por tanto será proponer una nueva forma de tratar los datos y extraer información útil de ellos mediante sistemas basados en redes neuronales convolucionales.

Aemas, este proyecto nos permite participar en el Andi Challenge mencionado anteriormente, el cual nos da la oportunidad de realizar un estudio competitivo y compartirlo con el resto de la comunidad científica interesada en este campo.

1.3 Objetivos

- Comprensión de la definición matemática del problema para su adaptación durante el proceso de resolución
- Desarrollo de un sistema automático de clasificación capaz de clasificar trayectorias reales en los 5 modelos estadísticos de movimiento anómalo explicados.
- Automatización del total de procesado y resolución del problema, para que los futuros usuarios tengan que hacer el mínimo trabajo posible.
- Conseguir una buena puntuación en la challenge para así poder poner en común nuestro método en conjunto con el resto de las participantes de esta.

2. Marco teórico

2.1 Membrana plasmática

Las células son la mínima unidad vital que compone todos los organismos, tanto unicelulares como pluricelulares. Dependiendo de la función que estas vayan a desarrollar presentan unas morfologías y componentes estructurales diferentes, pero todas ellas comparten una serie de características generales:

- Todas las células están rodeadas por una membrana muy delgada denominada membrana plasmática.
- Todas las células cuentan con un medio interno o medio intracelular denominado citoplasma, formado por sustancias orgánicas e inorgánicas, donde se encuentran sumergidas tanto las estructuras internas del citoesqueleto como los orgánulos celulares.
- Todas las células contienen materiales genéticos, en forma de ADN, ARN o ambas, necesario para llevar a cabo la síntesis de sustancias como para generar nuevas células (este material genético puede estar o no limitado dentro de un núcleo).

En este estudio nos centraremos en la membrana plasmática, una de las partes más interesantes de la célula debido a la gran cantidad de funciones que se llevan a cabo en esta, además de la compleja estructura que la forma. Por tanto, para hacernos una idea de la importancia de esta capa, hay que describirla tanto a nivel estructural como funcional.

2.1.1 Estructura

2.1.1.1 Estructura general

Esta sección de la célula, como indica su nombre, se trata de una capa, es decir, una superficie fina que rodea todo el interior de la célula (Figura 1). Esta membrana cuenta con un grosor de entre 7 y 10 nanómetros (dependiendo de la región), siendo por tanto imposible de ver mediante un microscopio óptico con resoluciones cercanas a los 200 nm. Por tanto, solo es visible mediante aparatos como el microscopio electrónico (Cooper, 2000; Castelle & Banfield, 2018).

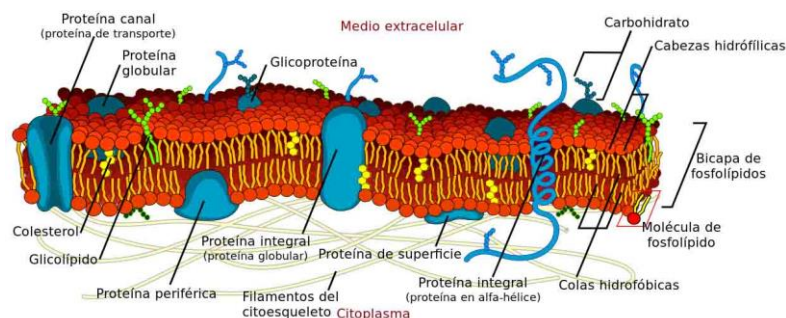


Figura 1: Membrana plasmática (Lizana Nuñez, 2015).

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Esta capa no se trata de una estructura estable, donde sus componentes estarían fijos en el espacio, sino que se trata de una superficie fluida (modelo del mosaico fluido), donde la gran mayoría de sus componentes están en continuo movimiento facilitando así las funciones que se llevan a cabo en esta.

En este mosaico fluido se encuentran gran cantidad de moléculas distintas, cada una con su función particular dentro del conjunto, las cuales presentaremos a continuación.

2.1.1.1.1 Lípidos de membrana

Se enmarcan en dos grandes grupos, los fosfolípidos y glicolípidos (Figura 2). Estas moléculas forman la llamada bicapa lipídica. Esta bicapa se trata de la mayor estructura de la capa, siendo aproximadamente un 50% de la masa total de esta. Esta capa se forma en forma de bicapa debido a las características de los lípidos que la componen, los cuales cuentan con dos dominios: apolar y polar. Debido a esta característica las moléculas tienden a interactuar con el medio extra e intracelular con su dominio polar y esconder su dominio apolar de estos. De esta forma los dominios apolares quedan enfrentados unos frente a otros formando una región libre de moléculas de H₂O y así una bicapa (Figura 3) (Van Meer, Voelker & Feigenson, 2008; Janmey & Kinnunen, 2006)

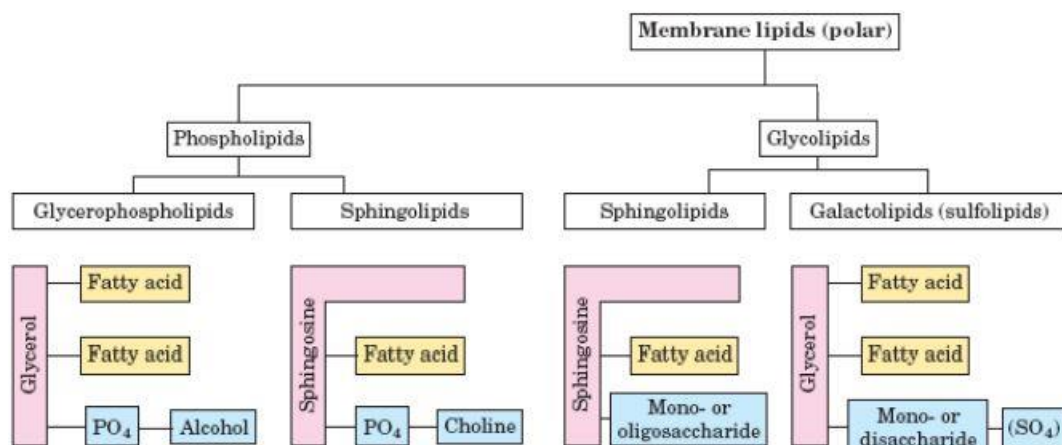


Figura 2: Tipos de lípidos de membrana divididos en dos grupos principales, fosfolípidos y glicolípidos. (Lison, 2018).

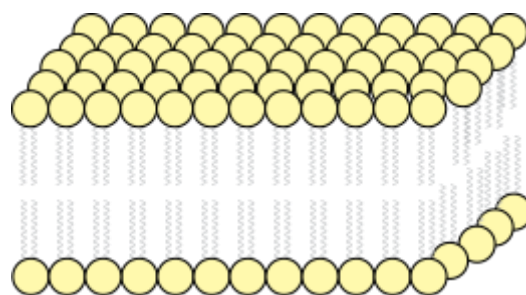


Figura 3: Bicapa lipídica, se observa como la parte apolar (colas en zig-zap) quedan enfrentadas entre sí, y la parte polar (bolitas) quedan en el exterior (Membrana Plasmática, 2020).

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Estos dos dominios están formados por diferentes componentes y estructuras los cuales les dotan de estas propiedades:

- **Dominio Polar** (cabeza polar): Esta parte está formado por una molécula de glicerol o de esfingosina, en el caso de los esfingolípidos (tanto en glicolípidos como fosfolípidos) aparece la molécula de esfingosina (2-amino-4-octadeceno-1,3-diol), por otra parte, en el caso de los glicerofosfolípidos y galactolípidos (sulfolípidos) encontramos una molécula de glicerol (propano-1,2,3-triol). Esta cabeza polar se puede unir tanto a un grupo fosfato unido a un grupo aminoalcohol o una molécula de colina (en el caso de los fosfolípidos) o un glúcido (mono u oligosacáridos) si se trata de un glicolípidos. La unión de estas cabezas a grupos tanto fosfatos como glúcidos determinara la función que llevaran a cabo, además, estos dominios se unen a los dominios apolares mediante enlaces éster en sus moléculas de glicerol o esfingosina.
- **Dominio apolar** (cola apolar): Estas colas se tratan de ácidos grasos, es decir, ácidos carboxílicos con cadena carbonatadas (siendo los ácidos carboxílicos moléculas polares debido a que presentan moléculas de oxígeno, que en medios acuosos adoptan una carga negativa, y la cadena carbonatada un grupo apolar, debido a la falta de grupos capaces de interactuar con las moléculas de agua, son cadenas de carbono e hidrogeno). Estas moléculas se pueden agrupar principalmente en dos grupos: saturados e insaturados. Siendo los ácidos grasos insaturados aquellos que presentan algún doble enlace en su cadena carbonatada, formando de esta manera doblesces (se desvía en cada doble enlace la linealidad de la cadena), y los saturados aquellos que no presentan ningún doble enlace y por tanto son totalmente rectos. Estos se unen con un enlace éster a las moléculas del dominio polar mediante su grupo acido carboxílico (dos ácidos grasos en el caso del glicerol y uno en el caso de la esfingosina).

2.1.1.1.2 Proteínas

Estas estructuras se encuentran dispersas alrededor de toda la membrana llevando a cabo principalmente tareas de transporte entre ambos medios que esta capa separa. Las proteínas, aunque cuentan con grandes diferencias entre ellas cuentan con una estructura y características generales comunes.

Las proteínas por definición tienen estructura de cadenas de aminoácidos (Figura 4), componentes primarios de los péptidos, los cuales están formados por una molécula central de carbono unida por sus cuatro radicales a un grupo carboxilo, un grupo amino, un grupo hidrogeno y una cadena lateral. La cadena lateral es la estructura que varía entre diferentes aminoácidos, siendo la cadena especifica la que determina el aminoácido del que se está hablando. (Branden & Tooze, 2012)

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

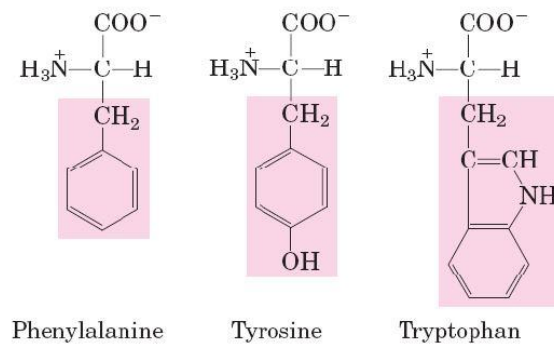


Figura 4: aminoácidos del grupo apolares aromáticos, varían su cadena lateral según el aminoácidos concreto (Lison, 2018).

Estos aminoácidos se unen entre si formando las dichas cadenas mediante enlaces amida del grupo carboxilo y el grupo amino.

Estas cadenas se disponen de una forma alfa hélice o beta-laminar para posteriormente plegarse sobre si mismas formando la estructura funcional de la proteína o estructura terciaria, si se unen varias de estas formando un nuevo conjunto funcional esta es denominada estructura cuaternaria (Branden & Tooze, 2012).

Dependiendo del orden de los aminoácidos que conformen la cadena y del largo de esta y de sus ramificaciones, esta proteína tendrá una estructura diferente, llevando a cabo así en la membrana diferentes funciones. Dependiendo de su disposición en la membrana podemos destacar principalmente dos tipos de proteínas (von Heijne, 2006):

- Proteínas integrales: cuentan con dominios que se extienden a lo largo de toda la membrana, es decir, la atraviesan por completo. Estas están relacionadas con el transporte de sustancias (tanto de forma activa o pasiva, dependiendo de si se requiere energía extra o funciona a favor del gradiente de concentración).
- Proteínas periféricas: Están unidos mediante interacciones electrostáticas o puentes de hidrogeno a los dominios hidrofílicos de las proteínas integrales o a las cabezas polares de los lípidos de membrana. Están relacionados con el reconocimiento de sustancias y señalización celular.

2.1.1.1.3 Glicoconjugados

Se tratan de cadenas de moléculas de glúcidos (tanto en forma de cadena como ramificadas) unidas a proteínas o lípidos de la membrana (glicoproteínas y glicolípidos). Las moléculas de glúcidos se tratan de cadenas de monosacáridos, siendo estas moléculas cíclicas con esqueleto central de carbono unidos a grupos hidrógenos e hidroxilos y cadenas laterales de propanol (del Campo, 1995).

Estas moléculas están especialmente relacionadas con funciones relativas a la inmunidad celular, comunicación celular y transporte y marcado de sustancias.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

2.1.1.1.4 Colesterol

Se trata de una molécula lipídica (del tipo esterol). En esta molécula se puede distinguir una cabeza polar constituida por el grupo hidroxilo y una cola o porción apolar formada por el carboxilo de núcleos condensados y los sustituyentes alifáticos (Figura 5).

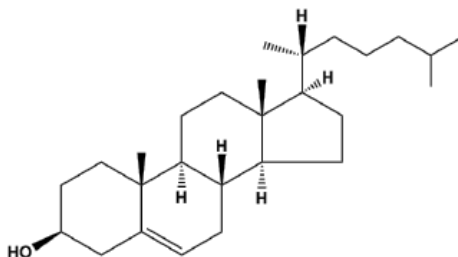


Figura 5: Molécula de colesterol completa, cada vertice corresponde con un átomo de carbono, y las aristas los enlaces covalentes entre estos (Lison, 2018).

Esta molécula se encuentra sumergida en la membrana plasmática, destacando su función de control de la fluidez, es decir, la presencia o ausencia de este componente en ciertas regiones de la membrana es lo que determina el movimiento general de los componentes de esta (Maldonado Saavedra, Ramírez Sánchez, García Sánchez, Ceballos Reyes & Méndez Bolaina, 2012).

2.1.1.2 Microdominios

Esta membrana no es uniforme en todas sus regiones, sino que presenta diferentes áreas o dominios debido a la concentración y disposición desigual de ciertas moléculas en el espacio. Estos dominios juegan papeles importantes en diferentes funciones de la membrana, aunque debido a las actuales tecnologías no se han conseguido caracterizar en cuanto a su tamaño, composición, abundancia y propiedades dinámicas.

La diferenciación entre diferentes microdominios está caracterizada por una serie de factores relativos a la presencia de ciertas moléculas en ellos, estos serían (Mukherjee, & Maxfield, 2004):

- **Presencia de colesterol:** La presencia de esta molécula, la cual interactúa directamente las cadenas de ácidos grasos y las moléculas de esfingosina de los lípidos de membrana, afectando así el orden de estos y a su fluidez, siendo directamente proporcional la presencia de colesterol en un área de la membrana con el orden de esta y por tanto su disminución de fluidez o movimiento de sus componentes. Actúa como aglutinante de los lípidos de membrana y da estabilidad.
- **Instauración de los lípidos de membrana:** Afecta directamente a la segregación de los lípidos en la membrana. La presencia de cadenas de ácidos grasos insaturados, es decir, con dobles enlaces en esta, dificultan la interacción de estas con las moléculas de colesterol, por tanto, dificultan la formación de dominios ordenados o estables juntamente con el colesterol.
- **Carga del dominio polar de los lípidos de los lípidos de membrana:** La carga de la cabeza polar de los lípidos afecta directamente al reclutamiento de proteínas en esa área.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

- **Presencia de proteínas de membrana:** En muchos casos, la polimerización de ciertas proteínas formando grandes grupos u oligómeros facilita la formación de microdominios de membrana estables, además de facilitar la unión de estos al citoesqueleto celular y las fibras de la matriz del medio extracelular.
- **Interacción con el citoesqueleto y matriz extracelular:** La adhesión de ciertas proteínas a las fibras de ambos esqueletos facilita la formación de microdominios estables y menos transitorios, restringiendo el movimiento de los lípidos que lo forman y están débilmente unidos a las proteínas.
- **Organización de las dos capas de la bicapa lipídico:** Siendo ambas capas de la membrana diferentes entre sí, en cuanto a presencia de ciertas proteínas, concentración y fluidez, se ha demostrado que las inhomogeneidades entre ambas capas influyen en la formación de ciertos microdominios debido a la conexión que existe entre ellas (seguramente debido a la presencia de ciertas proteínas integrales como las de la familia GPI). Siendo necesario el estudio de ambas capas de la bicapa de la misma área para determinar el microdominio que se está observando.

Podría decirse que hay tantos microdominios como tipos de proteínas se anclan a estos, pero principalmente distinguimos entre dos tipos de microdominios.

2.1.1.2.1 Caveolas

Microdominio caracterizado por ser una invaginación de la membrana de unos 55 nm aproximadamente (Figura 6). En muchos casos pueden asociarse varias caveolas dando lugar a regiones que presentan conjuntamente un gran número de invaginaciones. Estos microdominios o más bien sus mutaciones, según demuestran recientes estudios, están estrechamente relacionados con la señalización del óxido nítrico, el metabolismo lipídico y la regulación de la glucosa (Mukherjee, & Maxfield, 2004):.

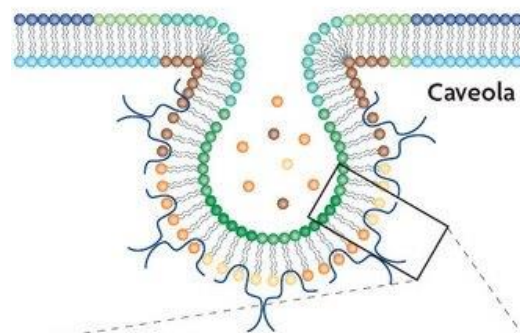


Figura 6: Caveola con sección de la membrana invaginada señalada (CajalseyGalileos, 2013)

2.1.1.2.2 Balsas Lipídicas (Lipid Rafts)

Microdominios muy estables formadas por la unión de colesterol y esfingolípidos con ácidos grasos saturados de cadena larga. Además de todos esto, cuentan con un gran número de proteínas ancladas a los lípidos, como la proteína GPI (Dietrich, Bagatolli, Volovyk, Thompson, Levi, Jacobson, & Gratton, 2001).

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Estas balsas potencian la generación de clusters de proteínas de señalización, atrayéndolas hacia estas zonas y haciendo que se unan entre si formando grandes estructuras poliméricas capaces de amplificar la señal a transducir al exterior o interior de la célula. Como sería el caso de las kinasas de la familia SRC, encargadas de funciones de síntesis proteica y crecimiento células, tendiendo a agruparse en estas balsas dando como resultado una mayor fosforilación y activación, amplificando así la señal (Mukherjee, & Maxfield, 2004).

Estos *lipid rafts* están estrechamente relacionados con la transducción de señales en procesos defensivos, conteniendo en ellas las proteínas encargadas de unirse con los antígenos y células T, contienen los TCR (receptores de células T). Estudios demuestran que la alteración de la composición de los *lipids rafts* afecta directamente a la señalización durante este proceso (Mukherjee, & Maxfield, 2004).

También están relacionados con la inmunidad a parásitos y trastornos alérgicos, provocando en ellos la unión de los receptores proteicos de la familia FcεRI, los cuales están estrechamente ligados a estos procesos, y que amplifican su señal en la unión con estos microdominios (Simons & Toomre, 2000).

Dependiendo de la disposición y concentración de las moléculas de estas balsas lipídicas, diferentes tipos de proteínas serán atraídas y clusterizadas en estos, por tanto, este microdominio supondrá un gran grupo de subdominios distintos y cada uno con una función específica (Figura 7).

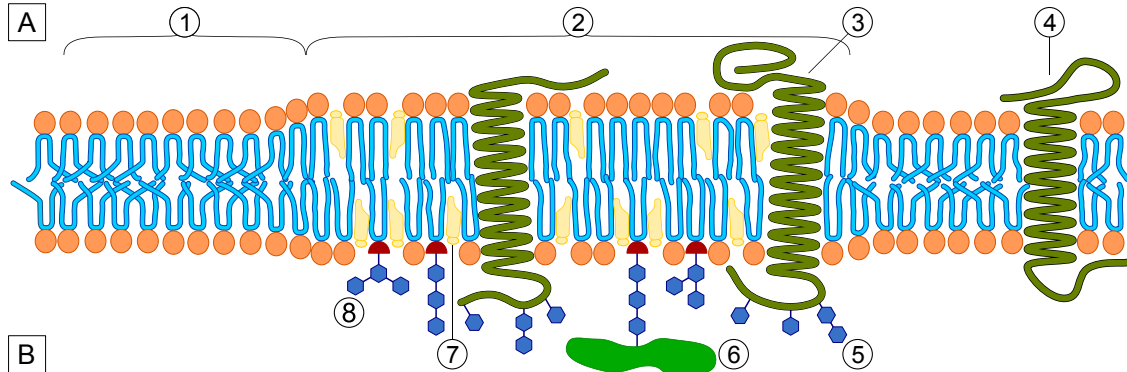


Figura 7: Corte transversal de la membrana con lipid raft (2), este presenta un mayor grosor que el resto debido a sus cadenas saturadas (Universidad de Alcalá (UAH), 2011).

2.1.1.3 Métodos de detección de microdominios

La presencia de estos microdominios por tanto es esencial para llevar a cabo las diferentes funciones de la célula, y el conocimiento y estudio del comportamiento y presencia de estos es esencial para el estudio del metabolismo y desarrollo celular.

Actualmente se emplean diferentes métodos para el reconocimiento de los diferentes microdominios a lo largo de la membrana, los cuales tienen como meta el mapeo total de la membrana

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

plasmática. Los métodos se pueden agrupar principalmente en tres grupos bien diferenciados (Mukherjee, & Maxfield, 2004).

2.1.1.3.1 Resistencia a Detergentes

La idea principal se basa en que aquellas capas de la membrana con mayor estabilidad, es decir, con un mayor orden de empaquetamiento (debido a la presencia de colesterol y saturaciones en los ácidos grasos), complicara el acceso de las moléculas iónicas detergentes a el núcleo hidrofóbico a la membrana (colas apolares y cuerpos de proteínas integrales). Por tanto, aquellos dominios más estables serán aquellos que no serán disueltos por parte de ciertas concentraciones de detergentes, aquellos con la misma estabilidad serán disueltos por la misma cantidad de detergente.

El principal problema de este método radica en la propia fluidez de la membrana, debido al movimiento continuo de esta, a la vez que se produce la solubilización de ciertas áreas de la membrana más desordenadas por parte del detergente, el resto de la membrana se reorganiza para poder cubrir esos huecos y de esta forma mantener la membrana a salvo de filtraciones.

2.1.1.3.2 Observación Directa

Estos métodos se basan en la observación mediante el ojo humano de ciertas características de la membrana. Dentro de este grupo encontramos dos métodos principalmente.

En primer lugar, la observación mediante microscopio óptico, empleando microscopia de campo laser amplio y confocal. Visualmente resulta imposible diferenciar entre los diferentes dominios, por tanto, se debe inducir la membrana a condiciones especiales. Para ello se deben emplear marcadores capaces de adherirse a los dominios ordenados (Dil-C16 o Dil-C18) y desordenados (C6-NDB-SM), además de reducir la presencia de colesterol. Empleando estos marcadores y con la ayuda de la reducción del colesterol, se pueden diferenciar las diferentes áreas que conforman la membrana.

En muchos casos, se emplea la unión de anticuerpos monovalentes marcados fluorescentemente para detectar microdominios, este es el caso de las proteínas unidas a los glicolípidos GPI (glicosilfosfatidilinositol), los cuales tienden a agruparse en balsas del tipo DRM y agruparse entre sí, formando así grandes polímeros que indican de forma indirecta que se encuentran en dominios ordenados de tipo balsa DRM.

En segundo lugar, el uso de las tecnologías *single particle tracking* (SPT), estas se basan en el marcaje directo de componentes de membrana mediante partículas de seguimiento (ej. Fluorocromos) para empleando cámaras de video de alta resolución y densidad de muestreo generar videos del movimiento de estas moléculas alrededor de la membrana. Explota la idea principal de que según el medio por el que se desplacen las moléculas, estas describirán un patrón de movimiento distinto, así pues, será posible conocer el medio mediante el movimiento de las moléculas sobre este (Manzo & Garcia-Parajo, 2015).

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Esta última tecnología está muy poco desarrollada ya que no cuenta con herramientas de análisis posterior del movimiento para la caracterización de microdomínios, ya que se ha demostrado una relación directa entre el movimiento de las moléculas y las características del medio en el que se desplazan. Esta tecnología es por tanto la que nos dedicaremos a explotar en este estudio.

2.1.1.3.3 Métodos Indirectos

Basadas en la medición de ciertas características de la membrana y de sus moléculas al recibir ciertos estímulos externos para así establecer relaciones directas con su estructura y organización espacial. Dentro de esta se emplean principalmente dos métodos.

En primer lugar, la transferencia de energía resonante por fluorescencia (FRET). Esta técnica emplea la radiación emitida por fluoróforos, que tienen una resolución de 3 a 10 nm, es decir, se pueden diferenciar fluoróforos que se encuentren a esta distancia entre sí como dos unidades diferentes. Siendo la extensión media de los microdominios de 100 nm, este método es bastante útil.

De normal se emplea la medición de la fluorescencia producto de ciertos procesos químicos. Por ejemplo, encontramos el caso de la despolarización de la fluoresceína (3',6'-dihidroxi-espíro[2-benzofuran-3,9'-xanten]-1-ona) en los receptores proteicos anclados a los glicolípidos GPI de membrana (Lichtman, & Conchello, 2005).

En segundo lugar, la resonancia por giro electrónico y anisotropía de fluorescencia. En este método se emplean marcadores como el nitróxido o el 3- (difenilhexatrienil) propanoil] -1- hexadecanoyl-sn-glicero-3-fosfo-colina (DPH-PC) en los lípidos de la membrana. El movimiento de translación lateral o rotación de estos afectan al espectro de resonancia del spin electrónico, alteración que puede ser captada para calcular la anisotropía de la membrana celular en sus diferentes áreas o regiones. Empleando el nivel de anisotropía se puede medir el nivel de ordenamiento de ciertas regiones, como se realiza en los sólidos para medir su nivel de cristalización (Mukherjee, & Maxfield, 2004).

2.1.2 Funciones

Esta capa cuenta con un gran número de funciones dentro del metabolismo y desarrollo celular, siendo una de las regiones de la célula que más variadas y distintas funciones lleva a cabo y, por tanto, un objeto de estudio indispensable para conocer más acerca de esta. Las funciones principales son las siguientes (Cooper, 2000):

- **Compartimentalización:** Mantiene el medio intracelular o citoplasma con todos los orgánulos que flotan en el separado del medio extracelular, actúa por tanto de límite o barrera entre medios.
- **Intercambio de moléculas:** La membrana plasmática, al ser el límite entre la región externa e interna de la célula, se ve obligada a actuar de mediador de sustancias entre ambas regiones. Esta capa puede transportar todo tipo de sustancias, las cuales dependiendo de sus características (masa, carga, concentración, procedencia, etc.) atravesar la membrana de una forma u otra. Las sustancias más pequeñas, solubles y neutras en cuanto a carga, podrán atravesar la membrana

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

sin interactuar con ninguna región especializada (directamente por la bicapa lipídica), en cuanto a las moléculas de mayor tamaño y con carga, tendrán que atravesar la membrana empleando canales proteicos activos y pasivos.

- **Reconocimiento y transducción de señales:** Mediante ciertas moléculas (proteínas y oligosacáridos), la membrana es capaz de detectar la presencia tanto en el medio extracelular como intracelular de ciertas sustancias, pudiendo así transducir estas señales a el medio extracelular para comunicarse con otras células vecinas u a el medio intracelular para enviar señales a ciertos orgánulos del citoplasma.
- **Reconocimiento:** Mediante las proteínas y oligosacáridos que conforman la superficie celular, los diferentes mecanismos del organismo son capaces de reconocer las células en cuanto a su procedencia y función. Esta característica es fundamental durante cualquier proceso inmunitario, ya que permite al organismo reconocer aquellos cuerpos celulares o moléculas que son exógenos al cuerpo, por tanto, desencadenar una respuesta inmunitaria dirigida a estos cuerpos.
- **Secreción y endocitosis:** Mediante invaginación o abultamientos de la membrana, la célula es capaz de absorber o expulsar grandes vesículas con sustancias de gran tamaño en su interior, moléculas que no podrían ser intercambiadas mediante canales proteicos estándar. Esta función también es clave en el proceso inmunitario, ya que permite a diferentes células defensivas del organismo tanto atrapar grandes cantidades de material extraño para disolverlo mediante vesículas ácidas en su interior, como expulsar sustancias corrosivas por tal de dañar agentes patógenos en ciertas áreas del organismo, que darían como resultado grandes infecciones.

2.2 Métodos de rastreo molecular

Recientemente se han desarrollado gran cantidad de métodos para medir las propiedades dinámicas de las moléculas en medios celulares. Todas ellas presentan ventajas en cuanto a una cierta configuración de resolución temporal y espacial, haciendo que para nuestro problema concreto tengamos que elegir aquella opción más favorable en cuanto a el compromiso que deseemos entre ambos tipos de resoluciones.

2.2.1 Tipos de métodos existentes

Entre estos destacan en primer lugar el FRAP (*fluorescent recovery after photobleaching*). En esta técnica se marcan una serie de moléculas en un área determinada o ROI con sustancias fluorescentes. Posteriormente, esa área es expuesta a un láser de alta potencia, que activa la fluorescencia de las moléculas ancladas. Después de un tiempo t , se mide la fluorescencia en el área marcada, de esta forma, se puede relacionar la cantidad de fluorescencia con el número de moléculas que han salido de la zona de estudio, y así caracterizar la difusión de estas (Manzo & Garcia-Parajo, 2015; Meddens, de Keijzer & Cambi, 2014)).

Este método tiene como desventaja que es incapaz de seguir trayectorias de partículas aisladas, solo es capaz de medir la difusión de estas en una región determinada.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Otros métodos empleados se tratan de la espectroscopia por correlación de fluorescencia (FCS) y la correlación por fluorescencia de imágenes (ICS). Estos métodos miden las variaciones de fluorescencia dentro de una región de estudio, estas variaciones pueden deberse tanto a entradas de nuevas moléculas en estas áreas, reacciones químicas o procesos de difusión. Posteriormente se establece una correlación con la fluorescencia media del área para determinar los cambios que se han producido dentro del área excitada (Magde, Elson, & Webb, 1974).

Estos métodos explicados anteriormente cuentan con una alta resolución temporal, es decir, una alta densidad de muestreo, capaz de detectar cambios rápidos entre reacciones y movimientos de las moléculas, pero cuentan con una reducida resolución espacial, debido principalmente a la difracción de la radiación emitida durante el proceso de fluorescencia. Con estos métodos se puede medir la media de las fluctuaciones dentro de un área, pero no es posible medir las diferencias entre áreas de la membrana al mismo tiempo.

Para solventar el problema de la resolución espacial se cuenta con dos técnicas e superresolución, el *stimulated emission depletion microscopy* (STED) y *near-field scanning optical microscopy* (NSOM). Ambos métodos emplean radiaciones de excitación por debajo de la difracción, de forma que son capaces de captar la topografía de la superficie molecular de forma muy exacta, con un muy bajo margen de error. Además de estos dos métodos, se están desarrollando otros similares basados en antenas y estructuras nanofotónicas, aunque aún se encuentran en fase de desarrollo. Estos métodos son capaces de medir la topografía espacial de forma muy precisa, aunque solo permiten hacer mediciones molécula a molécula, es decir, hacen barridos de la superficie empleando un solo punto de lectura, lo cual, en una superficie fluida y en continuo movimiento como es la membrana no resulta ventajoso (Betzig, Lewis, Harootunian, Isaacson & Kratschmer, 1986; Willig, Harke, Medda & Hell, 2007).

Como alternativa a estos métodos, destacamos el single particle tracking (SPT). Este método cuenta con un buen compromiso entre resolución espacial y temporal, además de empleándolo de forma correcta es capaz de hacer mediciones de forma simultánea de gran cantidad de moléculas a lo largo de toda la membrana. Esta técnica además es capaz de seguir a la misma partícula durante mucho tiempo, incluso minutos, para poder dibujar todos sus recorridos.

2.2.2 Single particle tracking (SPT)

Vistas las opciones que se disponen, y las ventajas de cada método, actualmente se opta por el método SPT, debido a su compromiso entre resolución espacial y temporal, además de su sencilla implementación respecto al resto de métodos más precisos.

2.2.2.1 Partículas de seguimiento

Las tecnologías SPT comenzaron empleando fluorocromos, marcando las moléculas a rastrear con sustancias fluorescentes, estas sustancias emiten radiación en el rango de 20 a 50 nm, lo cual dificulta

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

mucho obtener una resolución espacial favorable debido a la difracción, siendo mucho más útil emplear técnicas como el STED o NSOM.

La gran revolución de esta técnica empezó con la utilización de nanocristales semiconductores o *quantum dots* (QD). Estos puntos cuánticos son nanoestructuras (generalmente en forma de pirámide) sintetizados artificialmente en laboratorios a partir de decenas a cientos de átomos (dependiendo del tamaño que deseemos).

Estos nanocristales emiten radiación lumínica cuando son irradiados por un láser externo. Esta irradiación provoca la migración de electrones de la capa más baja de conducción a la última capa de valencia, cuando dejan de ser excitados estos vuelven a su posición expulsando fotones asociados a una longitud de onda concreta. Principalmente cuentan como ventaja que se requieren láseres de muy baja potencia, emitiendo radiaciones con longitudes de onda muy bajas (por debajo de la refracción), además, cada configuración espacial de estas nanoestructuras, es decir, la forma con la que han sido sintetizados, emitirá una radiación a una longitud de onda específica. Estas dos características permiten tanto hacer mediciones muy precisas de la posición de las moléculas marcadas, como evitar superposiciones de radiaciones de moléculas diferentes, pudiendo así hacer el seguimiento simultáneo de un gran número de estas. Además, debido al reducido tamaño de estos nanoconductores, se evita que su masa afecte al movimiento de las moléculas a las cuales está unido, además de reducir al máximo la dispersión de rayleigh, que atenúa la radiación de entrada a la sexta potencia del diámetro de la estructura que atraviesa (Alivisatos, 1996; Manzo & Garcia-Parajo, 2015).

2.2.2.2 Funcionamiento del proceso

Una vez entendido esto, es hora de hablar de las técnicas que emplea el SPT para hacer el seguimiento de las partículas marcadas con los QD. Estas partículas son rastreadas mediante una cámara de video que emplea una lente óptica (Manzo & Garcia-Parajo, 2015).

La resolución dependerá tanto de la longitud de onda de la radiación emitida por cada QD y la apertura del objetivo de la cámara. Cada QD, será captado por el objetivo como un anillo de intensidad, el cual es posible modelar como una gaussiana (Figura 8),

$$I(x, y) \approx I_0 \exp\left\{-\frac{(x - x_0)^2}{2w^2}\right\} \exp\left\{-\frac{(y - y_0)^2}{2w^2}\right\}.$$

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

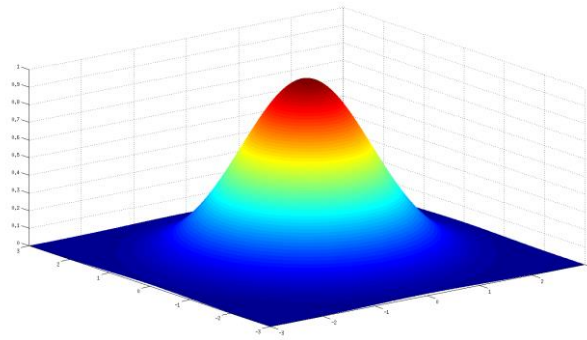


Figura 8: Función Gaussiana 3D

Partiendo de esta ecuación, donde I_0 es la intensidad del punto más alto de la gaussiana se extrae el parámetro FWHM (Full weight-half máximo), el cual es proporcional tanto a la longitud de onda (λ) como a la apertura del objetivo (NA)

$$FWHM = w\sqrt{8\log(2)} \approx \frac{\lambda}{2NA}. \quad (1)$$

Este efecto es la principal fuente de ruido, ya que no se podrán distinguir dos objetos entre sí que estén más cerca de la distancia que marca su FWHM. Este ruido es de tipo gaussiana, es decir, la probabilidad de encontrar la partícula en una posición concreta del aro lumínico viene dado por una gaussiana y con su punto máximo en el centro. Es un ruido modelable y a tener en cuenta para ciertos experimentos, ya que las partículas deberán de estar alejadas entre si con una cierta distancia que somos capaces de calcular.

Este ruido espacial, que afecta por tanto a la resolución espacial del objetivo, es producido por diferentes fuentes independientes entre sí. En primer lugar, la difracción del medio, los fotones son desviados de su dirección según se acercan al objetivo, mucho más según este alejado este. En segundo lugar, la anisotropía del medio, es decir, las emisiones de radiación fotónica no se distribuirán de forma homogénea por el medio, formando superficies esféricas con la misma intensidad a la misma distancia del núcleo irradiador, sino que habrá zonas a la misma distancia que reciban más intensidad lumínica.

Otro factor para tener en cuenta para la resolución espacial es la pixelación de la cámara, es decir, la cantidad de discretizaciones en el espacio que es capaz de representar. A mayor número de píxeles tenga el sistema de captación, mayor será la resolución espacial, dando un punto más concreto para la localización del centro de las gaussianas y sus límites.

Una vez marcadas las moléculas se realizan videos, es decir, tomas de imágenes sucesivamente y en cascada que duran varios minutos. La cantidad de tomas dependerá de la densidad de muestreo de la cámara, de normal estas están entre los 10 Hz las cámaras más convencionales y los 40000 Hz las más modernas y aun en desarrollo.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Los centroides de cada gaussiana tomada en las sucesivas imágenes del video del pueden unir para generar la trayectoria de cada molécula en el espacio, esta contara con tramos menos rectilíneos y por tanto será más realista según se aumente la densidad de muestreo. La resolución espacial por tanto viene determinada únicamente por la densidad de muestreo de la cámara que estemos empleando.

Para el estudio de la membrana celular se suelen emplear microscopios de fluorescencia de reflexión interna total (TIRF) (Figura 11). Esta modalidad permite excitar solo zonas superficiales de las muestras, lo cual es muy beneficioso a la hora de estudiar las membranas o capas más superiores de las muestras. El láser es proyectado de forma lateral hacia la muestra (luz de excitación), esta radiación excita ciertas moléculas selectivas en el área (Fluorocromos o nanopartículas como las QD). Este luz gracias al cristal de cuarzo sobre el que se dispone la célula es difractado totalmente, es decir, saldrá con el mismo ángulo de diagonal que como ha golpeado a la superficie celular. Este haz portara la radiación producida por la excitación de las moléculas selectivas y así será proyectado sobre un objetivo dispuesto en la trayectoria de difracción. En este objetivo se dispone de una cámara EMCCD, la cual distribuye el espectro electromagnético a lo largo de un plano discreto para poder determinar la posición de cada QD en cada toma (Fish, 2009).

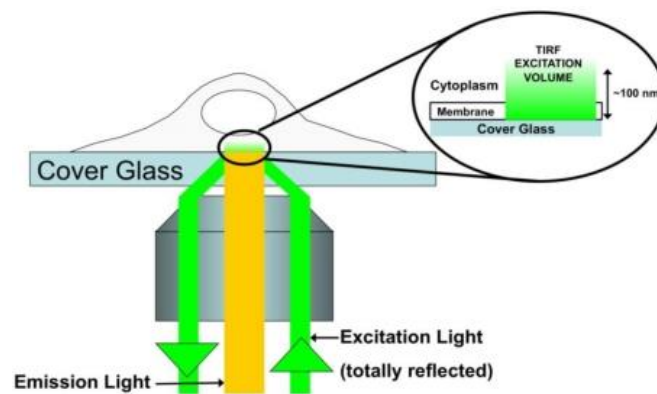


Figura 9: Microscopio TIRF (Fogarty, Zhang, Grigsby, Johnson, Chen, Mueller & Mansky, 2011).

Posteriormente, una vez obtenidas las imágenes sucesivas es necesario su preprocesamiento para obtener así las trayectorias finales. Este proceso sigue dos pasos sucesivos, la localización de los puntos en el espacio y la unión de los puntos para generar las trayectorias.

Para la localización de los puntos en el espacio se debe conocer la FWHM, es decir, las longitudes de onda con las que estamos trabajando y la apertura del objetivo. Con estos parámetros se puede formar la *Point Spread Function* (PSF) de las partículas. Esta función se trata de una gaussiana simétrica sobre el plano 2D, que su convolución con el aro de intensidades que devuelven los QD da la posición de la partícula, es decir, generando una función delta en el punto central de esta.

Debido a el error que produce la FWHM, la posición real de la partícula no puede ser determinada con exactitud, sino que se cometerá un leve error, aunque la posición con mayor probabilidad de encontrarla según la gaussiana será el centroide de esta (Figura 10).

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

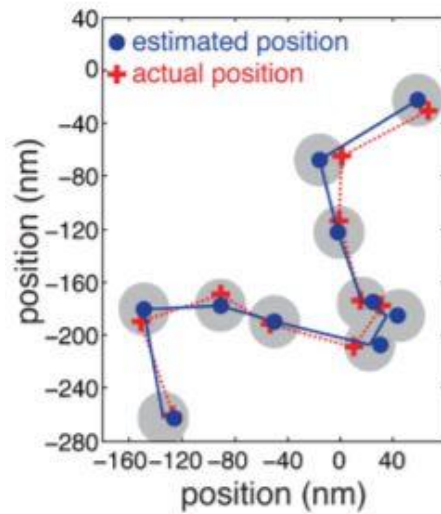


Figura 10: Posición real vs posición estimada (Manzo & Garcia-Parajo, 2015).

Para la unión de los sucesivos puntos se emplean algoritmos más sencillos, estos se basan únicamente en la unión de los puntos obtenidos en el paso anterior siguiendo la sucesión marcada por la toma de imágenes. En algunos casos puede haber imágenes donde no se detecte algún punto relativo a un QD específico, debido a problemas de difracción o de captación por el objetivo.

En algunos casos, debido a problemas de difracción o de captación del objetivo, no se detectan ciertos puntos de algún QD específico. Para solventar este problema se emplean algoritmos optimizados de interpolación. Estos algoritmos son capaces de generar puntos de forma artificial basándose en los puntos anteriores y posteriores a este, además de en la tendencia general de la función total, es muy importante por tanto ser capaces de describir la tendencia general de las trayectorias.

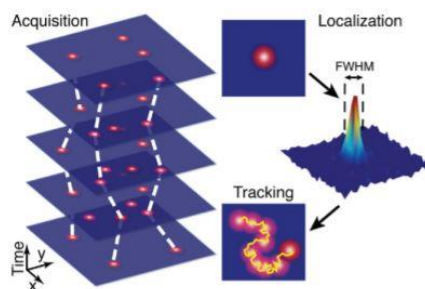


Figura 11: Proceso completo SPT (Manzo & Garcia-Parajo, 2015).

Una vez obtenidas las trayectorias finales debemos ser capaces de analizarlas, es decir, de extraer de ellas descriptores capaces de agruparlas en diferentes grupos o modelos. Esta tarea requiere un mayor conocimiento matemático y se comentará en el siguiente apartado.

2.3 Modelos difusivos

El movimiento browniano fue descrito por primera vez por Einstein y Smoluchowsky en 1905. Este movimiento es el que experimentan las partículas al colisionar e interactuar con otras dentro de un medio específico.

Estos movimientos son puramente estocásticos, es decir, los avances de las partículas en el medio son aleatorios y siguen una función de densidad de probabilidad (PDF) $P(x, t)$. Esta función nos indica la probabilidad que tendrá la partícula de encontrarse en el punto del espacio x en un tiempo concreto t desde el inicio de su movimiento. Se puede considerar como condición inicial la de la partícula en el origen $P(x = 0, t = 0) = 0$.

Las partículas que experimentan este tipo de movimiento están sujetas a la ecuación de la difusión

$$\frac{\partial P(x,t)}{\partial t} = D \frac{\partial^2 P(x,t)}{\partial x^2}. \quad (2)$$

En nuestro caso, al tratarse de partículas con coordenadas 2D, la expresión será la siguiente:

$$\frac{\partial P(\vec{r}, t)}{\partial t} = D \nabla^2 P(\vec{r}, t),$$

siendo D el coeficiente de difusión del medio, un valor que nos representa la facilidad o dificultad que tiene una partícula concreta para desplazarse en un medio específico. El vector \vec{r} nos indica la coordenada de la partícula y ∇ el vector operador diferencial del vector, es decir, la derivada de la función respecto a un vector.

El desplazamiento de las partículas será por tanto la variable aleatoria que registrará su probabilidad relativa según la función de densidad PDF. Cada instante temporal esta adoptará un valor según lo indique la función de densidad. Esta PDF se puede modelar como una distribución de valores en espacio, siendo en el caso de los espacios 2D una función 3D, con tres variables, las variables espaciales x e y indicadas por el vector \vec{r} , y la función temporal t .

Dependiendo de la forma de la PDF, la partícula tendrá un régimen de movimiento específico, por tanto, esta función determinará el tipo de movimiento de esta.

2.3.1 Random walks

Aunque en la práctica estos movimientos difusivos sean de carácter continuo, es difícil trabajar con estos de forma continua fuera del marco puramente teórico. Así pues, estos movimientos con tal de poder trabajar con ellos de forma computacional tendrán que ser de carácter discreto, es decir, contar cada trayectoria de cada partícula con un número finito de puntos o valores. Es por esto que definimos el concepto de random walk.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Sea un random walk una sucesión de puntos finita que representa el avance de una trayectoria en el espacio. En esta sucesión está caracterizada por tanto por una variable temporal y una espacial. La variable temporal es constante, ya que la separación entre dos puntos sucesivos de la sucesión es siempre la misma (Incrementos de t constantes), en cambio, los valores que figuran en la sucesión hacen referencia a la variable espacial, los cuales siguen una distribución dada por una función de densidad de probabilidad, pueden ser dependientes o independientes entre sí (Klafter & Sokolov, 2011),

$$X(t) = (x_1, x_2, x_3, \dots, x_n). \quad (3)$$

2.3.2 Mean square displacement

Dentro de la amplia gama de descriptores que se pueden extraer de estos RW, uno de los más interesantes y que será muy útil para caracterizar diferentes regímenes de movimiento se trata del *Mean square Displacement* (MSD). Esta variable describe la media de los desplazamientos desde el punto inicial hasta el final del conjunto de trayectorias empleadas, empleando todos los intervalos de tiempo deseados (lapsos de tiempo). Es muy sencillo de implementar ya que los intervalos de tiempo entre sucesivos valores son constantes (Klafter & Sokolov, 2011; Muñoz-Gil, Lewestein & Garcia-March, 2020; Guitou & Garcia-March, 2020)

$$MSD(t) = \langle (X(t) - X(0))^2 \rangle = \frac{1}{M} \sum_{i=1}^M (X_i(t) - X_i(0))^2. \quad (4)$$

Ya que trabajaremos con trayectorias en 2D, ampliaremos este descriptor para puntos con coordenadas en x e y , además de considerar el punto $X(0)$ como la coordenada $(0,0)$, por tanto, $\langle (X(t) - X(0))^2 \rangle = \langle X(t)^2 \rangle$. La ecuación para 2D quedara de la siguiente forma:

$$MSD(t) = \langle X(t)^2 \rangle = \langle x(t)^2 \rangle + \langle y(t)^2 \rangle, \quad (5)$$

mediante estas ecuaciones se extraerá una gráfica como la figura 12.

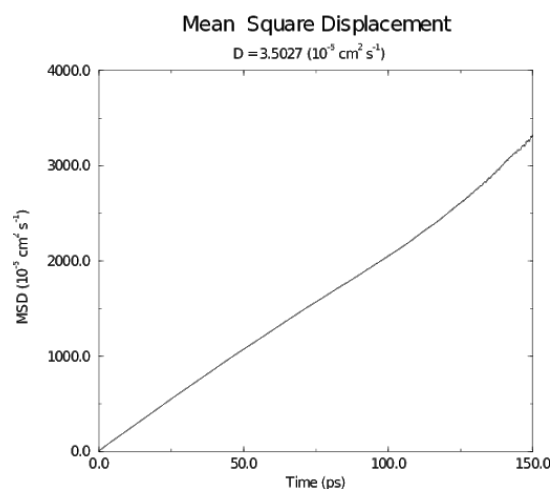


Figura 12: Gráfica MSD, empleando un coeficiente de difusión de 3.5027 y unas 100 trayectorias en total.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

En muchas ocasiones, es recomendable emplear un nuevo parametro, el tMSD. Este promedio nos permite analizar trayectorias individuales en vez de hacer promedios de grandes conjuntos.

El tMSD se calcula sobre una sola trayectoria empleando t lapsos diferentes. Es decir, no toma como punto de referencia el origen de la trayectoria, sino que toma los puntos separados según el t lapso que empleemos haciendo desplazamientos sucesivos de una sola unidad,

$$tMSD(t) = \overline{\delta^2(t)} = \frac{1}{N-n+1} \sum_{i=0}^{N-n} (X_{i+n} - X_i)^2 . \quad (6)$$

Por ejemplo, tomando una trayectoria de 10 puntos, queremos realizar la gráfica correspondiente al instante 2, es decir al t lapso=2. Se calcularía de la siguiente forma:

$$MSD(t_2) = \frac{1}{9} \sum_{i=0}^8 (X_{i+2} - X_i)^2 = \frac{1}{10} [X_2^2 + (X_3 - X_1)^2 + \dots + (X_{10} - X_8)^2] . \quad (7)$$

El uso indiscriminado de tMSD y MSD se cumple en los casos donde las trayectorias forman señales ergódicas. La ergodicidad implica que el promedio temporal de cualquier región de la señal es igual al promedio total de la señal, es decir, la media en cualquier segmento de la señal sea más largo o corto, sera igual a la media de la señal entera. Esto gráficamente quiere decir, que se cumplirá para aquellas trayectorias donde su grafica de MSD se presente lo más similar posible a una recta (debido a la estocástica siempre presentará ligeras irregularidades).

El empleo del tMSD es debido principalmente a que en la mayoría de los experimentos de SPT es muy complicado obtener una gran cantidad de trayectorias de forma simultánea del mismo modelo de movimiento, es decir, dentro de una misma área homogénea, y, por tanto, podría producirse un promedio erróneo.

En el caso de contar con trayectorias no ergodicas se romperá esta equivalencia entre MSD y tMSD, haciendo que la única forma de obtener el MSD sea mediante el promedio de muchas partículas que cuenten con el mismo régimen de movimiento.

También es posible el cálculo a nivel teórico del MSD si conocemos la función de densidad de probabilidad que rige nuestra trayectoria (en los casos que esta cuente con una sola PDF constante). Siendo en estos casos el MSD (para trayectorias en 2D), igual a la integral en todo el espacio del producto de la variable aleatoria al cuadrado y la PDF (Muñoz-Gil, Lewestein & Garcia-March, 2020),

$$\langle x^2(t) \rangle = \int_{-\infty}^{+\infty} x^2 P(x, t) dx . \quad (8)$$

Por ejemplo, si contamos con una distribución de trayectorias con una PDF gaussiana en el espacio

$$P(x, t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) , \quad (9)$$

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

siendo D el coeficiente de difusión y x el punto del que se estudia su probabilidad en el tiempo t . Su MSD será el siguiente:

$$\langle x^2(t) \rangle = \int_{-\infty}^{+\infty} x^2 P(x, t) dx = 2Dt. \quad (10)$$

En nuestro caso, contando con trayectorias en 2D, el MSD será $\langle \vec{r}^2(t) \rangle = 4Dt$.

2.3.3 Tipos de difusión

Una vez entendido este parámetro podemos diferenciar principalmente entre 4 tipos de movimientos según la forma de la gráfica de su MSD. Estos se tratan de la difusión normal, difusión dirigida, difusión confinada y difusión anómala. En el caso de la difusión anómala puede presentar diferentes graficas de MSD, tanto subdifusivas, superdifusivas o lineales con grandes niveles de ruido (Guitou & Garcia-March, 2020; Wagner, Kroll, Haramagatti, Lipinski & Wiemann, 2017).

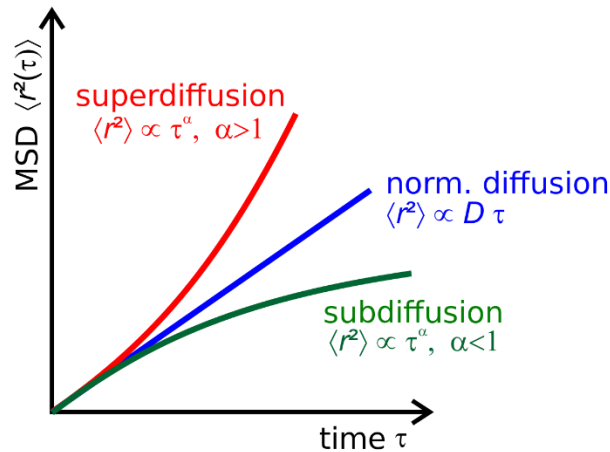


Figura 13: Tipos de difusión según MSD, con vector r como coordenada espacial 2D y τ como la variable temporal

2.3.3.1 Difusión Normal

En este modelo el movimiento es totalmente aleatorio en todas las direcciones del espacio posibles, es decir, presentará una PDF gaussiana (ecuación 9) y por tanto su MSD será lineal y como en el anterior ejemplo.

Este modelo de difusión se manifiesta principalmente en el movimiento de los lípidos de membrana en las regiones homogéneas de la membrana donde su movimiento no se ve afectado por ningún subdominio específico.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

2.3.3.2 Difusión Confinada

Se trata de un régimen de difusión normal el cual es modelado mediante un parámetro B que caracteriza cuanto esta confinada la trayectoria dentro de su área restringida definida como un área rectangular

$$A_{rect} \approx 4Dt ,$$

definimos así el área de la elipse más pequeña posible encerrada dentro del rectángulo

$$A_{ell} \approx \frac{\pi}{4} A_{rect} = \pi Dt ,$$

por tanto, el parámetro B en relación con el área de la elipse de confinamiento y el radio r donde se podría encontrar la partícula,

$$B = \frac{A_{ell}}{\pi r^2} = \frac{Dt}{r^2} .$$

Conjuntamente estas expresiones con la difusión normal se obtiene la siguiente función de MSD

$$\langle r^2(t) \rangle \approx r_c^2 \left[1 - A_1 \exp \left(- \frac{4A_2 Dt}{r_c^2} \right) \right] ,$$

empleando r_c como radio de la elipse para diferenciarlos del vector r de coordenadas (x,y) .

Esta debido a su carácter exponencial no es estrictamente ergódica, aunque debido a el leve carácter exponencial que presenta puede analizarse como tal.

Este régimen difusivo se presenta principalmente en las regiones donde los glicolípidos y lipoproteínas se encuentran sujetas al citoesqueleto o a la matriz extracelular, haciendo que su movimiento quede restringido dentro de pequeñas áreas (generalmente elípticas).

2.3.3.3 Difusión dirigida

Estos regímenes de movimiento se tratan de difusiones con una dirección de desplazamiento preferente en cada instante, mostrando graficas de desplazamiento siguiendo un patrón constante. En estos movimientos la dirección ya no es gaussiana, sino que sigue una función, como la longitud entre saltos de los intervalos temporales. Estos modelos están gobernados por el parámetro R .

Se forman por la suma de dos trayectorias simultáneamente, es decir, combinando instante a instante ambas. En primer lugar, se emplea una trayectoria generada con un modelo de difusión normal que aporta el componente de desplazamiento entre saltos sucesivos. En segundo lugar, se

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

emplea una trayectoria activa pura, es decir, una trayectoria cuyos desplazamientos se deben al producto de los intervalos de tiempo y la velocidad de cambio de dirección: $d_i = v\Delta t$ (se emplea una velocidad constante).

El resultado de combinar ambas da como resultado un parámetro R de la forma

$$R = \frac{v^2 T}{4D},$$

y finalmente una expresión de MSD

$$\langle r^2(t) \rangle = 4Dt + (vt)^2.$$

Este caso es análogo al anterior, debido a su carácter exponencial no es estrictamente regordido, pero al ser este muy débil se puede considerar así en la práctica.

Este movimiento se presenta principalmente en las proteínas de membrana que migran entre microdominios, atraídos por estos. Estas al ser liberadas de un microdominio específico se dirigen a otro similar que no haya sido disuelto, y para ello emplean una dirección preferente, que es aquella que resulta de generar un camino más corto entre ambos puntos, la rectilínea (manteniendo ciertas desviaciones debido a la estocástica del movimiento).

2.3.3.4 Difusión anómala

Este modelo de movimiento cuenta con diferentes comportamientos, pudiendo dibujar diferentes gráficas de MSD según el tipo de anomalía que presente. Por tanto, es imposible definir un solo modelo de movimiento anómalo, sino que dentro de este grupo contaremos con muchos modelos. Como característica principal presentan que la mayoría de sus MSD estarán modelados mediante un coeficiente anómalo α , que se modelara generalmente de la forma $MSD = K_\alpha t^\alpha$, siendo K_α una expresión dependiente de α .

En la mayoría de los modelos de este grupo de difusión no se cumple el teorema de la ergodicidad, haciendo por tanto que no se cumpla la igualdad entre el tMSD y MSD (Muñoz-Gil, Lewestein & Garcia-March, 2020).

$$\overline{\delta^2(\Delta)} \neq \langle x^2(\Delta) \rangle_N$$

De esta forma, resulta imposible analizar el MSD de estos modelos con solo una o un número reducido de trayectorias, ya que no podemos generar un tMSD ni un MSD fiable. Por tanto, presentará un reto la clasificación de estos, ya que no se podrá emplear este parámetro.

Además de todo esto, la variación del coeficiente α puede suponer grandes cambios en el movimiento resultante. Dentro de un mismo modelo, dos trayectorias generadas con dos alfas muy

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

alejados entre sí pueden dar como resultado dos trayectorias prácticamente irreconocibles entre sí, llegando a pensar en muchos casos que se tratan de trayectoria distintas.

Para el análisis de los movimientos anómalos emplearemos los modelos que se manifiestan más en el ámbito biológico, es decir, los que describen mejor ciertos tipos de movimientos ocurridos en la membrana celular. Estos modelos están relacionados con el comportamiento de los lípidos de membrana y proteínas en microdominios específicos de la membrana celular.

Describimos en este apartado 5 modelos que se ha demostrado su prevalencia en entornos biológicos. Además, estos son los que se estudiarán en el trabajo y en el *challenge*.

2.3.3.4.1 Continuous-time Random walks (CTRW)

En este modelo la partícula se mantiene inmóvil durante un determinado tiempo para posteriormente realizar un salto a otra zona del espacio más alejada. Los tiempos de espera entre saltos y el desplazamiento entre saltos son las dos variables que lo definen y son además independientes entre sí, son variables a su vez estocásticas. Este tipo de movimientos se manifiesta por parte de los lípidos dentro de microdominios de señalización como las balsas lipídicas (Scher & Montroll, 1975; Muñoz-Gil, Garcia-March, Manzo, Martín-Guerrero & Lewenstein, 2019).

En primer lugar, hay que definir una PDF que nos indique la probabilidad de una partícula de permanecer inmóvil durante un tiempo t

$$\Psi(t) = \int_t^\infty \psi(t') dt' = 1 - \int_0^t \psi(t') dt', \quad (11)$$

la $\psi(t)$ es la distribución de tiempos de espera. Ahora hay que definir una nueva PDF que nos indique la probabilidad de que la partícula realice n pasos (siendo aquellos momentos donde permanece inmóvil un solo paso) en un tiempo t . Para ello combinamos la distribución de tiempos de espera y la PDF de tiempo de espera

$$\chi_n(t) = \int_0^t \psi_n(t) \Psi(t - \tau) d\tau. \quad (12)$$

Aplicamos la transformación de Laplace a esta expresión para futuros cálculos, quedando de la forma

$$\chi_n(s) = \psi^n(s) \frac{1 - \psi(s)}{s}. \quad (13)$$

Mediante esta expresión (en el dominio del tiempo) podemos definir una nueva PDF que nos muestre la probabilidad de encontrar la partícula en una posición x en un tiempo t

$$P(x, t) = \sum_n^\infty P_n(x) \chi_n(t), \quad (14)$$

siendo $P_n(x)$ la probabilidad de alcanzar la posición x en n pasos.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

Implicamos a continuación la transformada de Fourier-Laplace a la expresión. Siendo Fourier para la variable espacial y Laplace para la temporal. Quedando de a forma

$$P(k, s) = \frac{1-\psi(s)}{s} \frac{1}{1-\lambda(k)\psi(s)}. \quad (15)$$

Una vez obtenida esta expresión podemos proceder a calcular la expresión general del MSD para este tipo de movimientos. Siendo k la variable espacial en el dominio de Fourier y s la variable temporal en el dominio de Poisson.

Para este cálculo emplearemos los momentos de desplazamiento de la PDF. Siendo el segundo momento de desplazamiento de una PDF de este tipo el MSD. Estos momentos están definidos en el dominio de Fourier-Laplace de la siguiente forma:

$$M_n(s) = (-i)^n \frac{d^n P(k,s)}{dk^n} \Big|_{k=0}. \quad (16)$$

El resultado de generar este momento tendrá la forma

$$M_2(s) = \langle x^2(s) \rangle = -\frac{d^2 P(k,s)}{dk^2} = \frac{\Psi(s)}{s(1-\Psi(s))} \langle l^2 \rangle + \frac{2\Psi^2(s)}{s(1-\Psi(s))} \langle l \rangle, \quad (17)$$

siendo $\langle l^2 \rangle$ la varianza y $\langle l \rangle$ la media de la PDF. Suponiendo que la distribución en el espacio es gaussiana, la media de la distribución pasaría a ser 0, $\langle l \rangle = 0$, eliminando así el segundo termino de la expresión y quedando de la siguiente forma

$$\langle x(s)^2 \rangle = \frac{\Psi(s)}{s(1-\Psi(s))} \langle l^2 \rangle. \quad (18)$$

Esta expresión será por tanto la expresión general del MSD, pero tomará diferentes formas según la función de distribución de tiempos de espera que empleemos. Destacamos principalmente las dos distribuciones más comunes dentro de este modelo, la de Poisson y la *heavy-tailed* (Guitou & Garcia-March, 2020).

Antes de explicar ambas definimos el siguiente parámetro que nos indicara el número medio de pasos realizados en un tiempo t . Además, pasamos esta expresión al dominio de Laplace dando como resultado

$$\langle n(t) \rangle = \sum_{n=0}^{\infty} n \chi_n(t) \xrightarrow{\mathcal{L}} \langle n(s) \rangle = \frac{1-\Psi(s)}{s} \sum_{n=0}^{\infty} n \Psi^n(s), \quad (19)$$

empleando la propiedad $\sum_{n=0}^{\infty} n a^n = \frac{1}{1-a^2}$ la expresión resultante será de la forma

$$\langle n(s) \rangle = \frac{\Psi(s)}{s(1-\Psi(s))}, \quad (20)$$

por tanto, quedara demostrada la relación

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

$$\langle x(s)^2 \rangle = \langle n(s) \rangle \langle l^2 \rangle \Leftrightarrow \langle x(t)^2 \rangle = \langle n(t) \rangle \langle l^2 \rangle . \quad (21)$$

Esta igualdad del dominio de Laplace y el del tiempo en el MSD nos será muy útil para simplificar futuros cálculos.

Heavy-tailed

Consideramos una PDF de la siguiente forma con un alfa en el dominio $0 < \alpha < 1$. Esta PDF la expresamos en el dominio de Laplace como

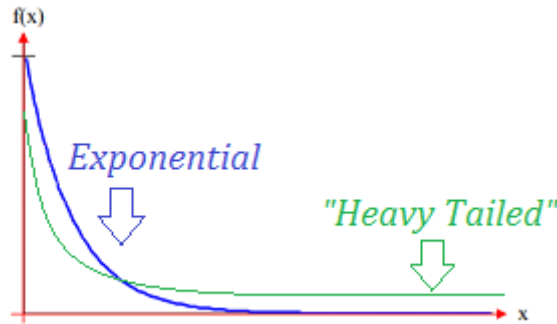


Figura 14: Distribucion exponencial y heavy-tailed, siendo x la variables que modela la función $f(x)$ en el espacio (Glen, 2016).

$$\Psi(t) \approx \frac{\alpha}{\Gamma(1-\alpha)} \frac{\tau^\alpha}{t^{1+\alpha}} \xrightarrow{\mathcal{L}} \Psi(s) = 1 - \tau^\alpha s^\alpha . \quad (22)$$

Empleando el teorema Tauberiano, el cual nos indica que para una $t > 0$ y para una expresión de la forma $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$, se satisface la propiedad $\Gamma(1+z) = z\Gamma(z)$. De esta forma la transformada de Laplace de $\Psi(t)$ puede emplearse para resolver el teorema Tauberiano. Conocemos esto de la ecuación (20). Dando como resultado

$$\langle n(s) \rangle = \frac{1}{\tau^\alpha s^{\alpha+1}} \xrightarrow{\mathcal{L}^{-1}} \langle n(t) \rangle \frac{1}{\Gamma(\alpha+1)} \frac{t^\alpha}{\tau^\alpha} , \quad (23)$$

empleando la transformada de Laplace inversa y la ecuación (21) obtenemos

$$\langle x^2(t) \rangle = \frac{1}{\Gamma(\alpha+1)} \frac{\langle l^2 \rangle}{\tau^2} t^\alpha , \quad (24)$$

aplicando la propiedad tauberiana de las funciones Γ , obtenemos

$$\langle x^2(t) \rangle = \frac{1}{\alpha \Gamma(\alpha)} \frac{\langle l^2 \rangle}{\tau^\alpha} t^\alpha , \quad (25)$$

teniendo en cuenta la expresión del coeficiente de difusión generalizado es

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

$$K_\alpha = \frac{1}{2\alpha\Gamma(\alpha)} \frac{\langle l^2 \rangle}{\tau^2}, \quad (26)$$

la expresión final quedara de la forma

$$\langle x^2(t) \rangle = 2K_\alpha t^\alpha, \quad (27)$$

siendo la expresión final del MSD en 2D

$$MSD = 4K_\alpha t^\alpha \quad 0 < \alpha < 1, \quad (28)$$

la expresión resultante se trata de una exponencial, por tanto, tendrá un comportamiento superdifusivo.

Poisson

Consideramos la siguiente PDF para los tiempos de espera

$$\Psi(t) = \lambda e^{-\lambda t}. \quad (29)$$

Aplicando la transformada de Laplace.

$$\mathcal{L}(\Psi(t)) = \Psi(s) = \frac{\lambda}{s + \lambda}, \quad (30)$$

de la ecuación (20) podemos deducir la expresion

$$\langle n(s) \rangle = \frac{\lambda}{s^2} \Leftrightarrow \langle n(t) \rangle = \lambda t, \quad (31)$$

así, de la ecuación (21) obtenemos

$$\langle x^2(t) \rangle = \lambda t \langle l^2 \rangle. \quad (32)$$

Teniendo en cuenta que $D = \lambda \langle l^2 \rangle$ el resultado será

$$\langle x^2(t) \rangle = Dt, \quad (33)$$

siendo en 2D

$$MSD = 2Dt. \quad (34)$$

Se trata por tanto de una expresión lineal, por tanto, ergódica.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES



Figura 15: Ejemplo de CTRW de 1000 pasos

2.3.3.4.2 Fractional Brownian Motion (FBM)

La difusión *Fractional Brownian Motion* (FBM) se trata de un proceso Gaussiano $B_H(t)$ definido en $[0, T]$ que empieza en 0. Se define como (Guitou & Garcia-March, 2020)

$$B_H(t) = \frac{1}{\Gamma(H+\frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}} dB(s), \quad (35)$$

donde el coeficiente de Hurst (H) está relacionado con el parámetro alfa por la siguiente relación:

$$\alpha = 2H .$$

Este modelo es muy similar a un movimiento browniano normal, es decir, a una difusión normal. La principal diferencia se encuentra en la correlación de los pasos, en el movimiento normal estos son independientes unos de otros, en el FBM no, todos los pasos están correlacionados entre sí de forma directa según marca la primera ecuación.

Dependiendo del valor de H el FBM tendrá un régimen diferente. Si $H=1/2$ significara que $\alpha = 1$, por tanto, corresponderá a una difusión normal. Si $H>1/2$ será un modelo superdifusivo debido a que $\alpha>1$. En cambio, si $H<1/2$ corresponderá a un modelo subdifusivo ya que $\alpha<1$.

Por ejemplo, en el caso superdifusivo la expresión del MSD tendrá la forma

$$MSD = 2dKt^\alpha \quad \alpha > 1 .$$

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Este modelo difusivo está estrechamente ligado a todas las moléculas de membrana que quedan ligadas a mRNA (ARN mensajero), es decir, encargadas de la síntesis de proteínas en microdominios especializados ribosómicos.

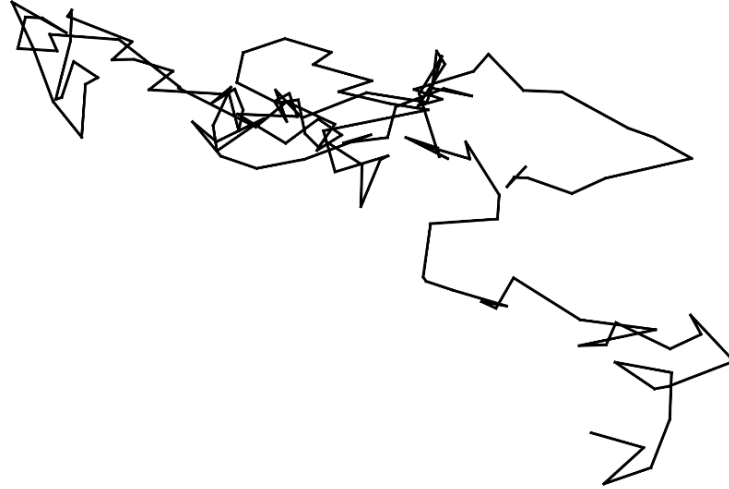


Figura 16: Ejemplo de FBM de 1000 pasos

2.3.3.4.3 Annealed transit time model (ATTM)

Este modelo llamado *Annealed transit time model* (ATTM), presenta grandes similitudes con el modelo CTRW ya explicado, pudiendo considerarse un tipo de CTRW. La forma de moverse y las ecuaciones que lo rigen son prácticamente las mismas salvando pequeños detalles que no serán necesarios comentar para comprender el funcionamiento de este modelo. Solo variará la difusión. Debido a su similitud, este se manifiesta en entornos similares a el CTRW, es decir, microdominios *lipid rafts*, modelando aquellos más inhomogéneos e inestables, debido al cambio del coeficiente de difusión del medio (Muñoz-Gil, Garcia-March, Manzo, Martín-Guerrero & Lewenstein, 2019; Massignan, Manzo, Torreno-Pina, García-Parajo, Lewenstein, & Lapeyre, 2014).

Con variar la difusión nos referimos a el coeficiente de difusión. Es decir, en este modelo la partícula se moverá durante un tiempo t con un coeficiente de difusión D , formando pares $(\tau|D)$. Asumimos que estos pares de tiempo y difusión se distribuyen y pueden modelarse mediante una PDF de la forma

$$P_{D,\tau}(D, \tau) = P_D(D)P_\tau(\tau|D), \quad (36)$$

siendo $P_D(D) \sim D^{\alpha-1}$ cuando $\alpha > 0$.

El resultado de las MSD será similar al de los modelos CTRW, dependiendo de la PDF que tomemos para los tiempos de espera (*Heavy-tailed* o de Poisson), y teniendo en cuenta el cambio del coeficiente de difusión.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Este modelo puede ser expresado mediante diferentes formas, dependiendo de cómo se desee regular el cambio de coeficiente de difusión, entre ellas está el ARM, basado en el radio, es decir, los cambios de D se harán según la partícula salgo de una determinada área.

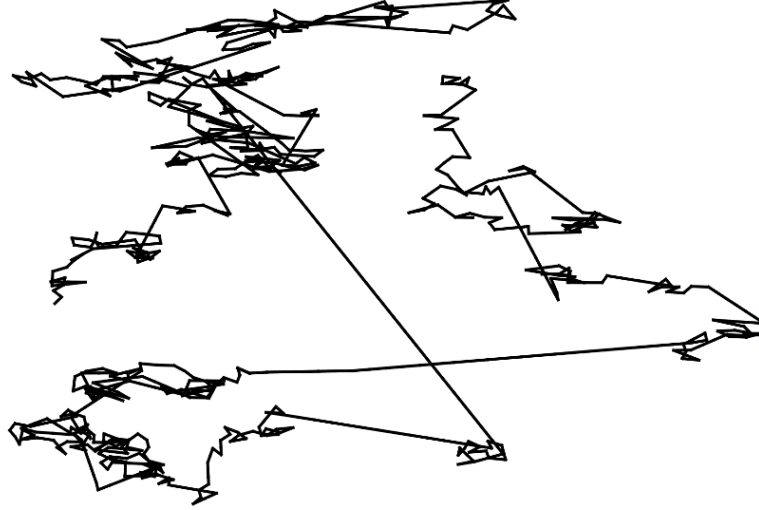


Figura 17: Ejemplo de ATTM de 1000 pasos

2.3.3.4.4 Scaled Brownian Motion (SBM)

El modelo *Scaled Brownian Motion* (SBM) comparte grandes similitudes con el modelo difusivo normal, al igual que el FBM. Este modelo parte de la PDF gaussiana de los modelos difusivos normales, solo que a la variable tiempo se le aplica la transformación no-lineal (Lim & Muniandy, 2002)

$$t \rightarrow t_* = t^\alpha \quad 0 < \alpha < 2, \quad (37)$$

pasando así a ser una variante dependiente del coeficiente anómalo alfa. Perteneciendo esta t a una PDF gaussiana de la forma

$$P(x, t) = \frac{1}{\sqrt{4D\pi t}} \exp\left[-\frac{x^2}{4Dt}\right]. \quad (38)$$

Como ya se ha explicado anteriormente, todas las partículas difusivas deben cumplir la ecuación de Einstein, que en este caso tendría la forma

$$\frac{\partial P(x, t_*)}{\partial t_*} = D \frac{\partial^2 P(x, t_*)}{\partial x^2}, \quad (39)$$

la solución de esta nueva ecuación será dada por la expresión

$$P(x, t_*) = \frac{1}{\sqrt{4D\pi t_*}} \exp\left[-\frac{x^2}{4Dt_*}\right] = \frac{1}{\sqrt{4D\pi t^\alpha}} \exp\left[-\frac{x^2}{4Dt^\alpha}\right] \equiv P_*(x, t), \quad (40)$$

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

a la cual aplicando la propiedad

$$\frac{d}{dt} = \frac{dt_*}{dt} \frac{d}{dt_*} = \alpha t^{\alpha-1} \frac{d}{dt_*}, \quad (41)$$

deducimos que la nueva ecuación puede ser escrita como:

$$\frac{dP_*(x,t)}{dt} = \alpha D t^{\alpha-1} \frac{d^2 P_*(x,t)}{dx^2} = D_* \frac{d^2 P_*(x,t)}{dx^2}. \quad (42)$$

Considerando que el nuevo coeficiente de difusión también crecerá en función del coeficiente anómalo y su valor inicial $D_*(t) = \alpha D t^{\alpha-1}$. Además, considerando las transformación escalares: $t \rightarrow bt$ y $x \rightarrow b^{\alpha/2}x$, la PDF podría ser escrita como

$$P_*\left(b^{\alpha/2}x, bt\right) = b^{\alpha/2} P_*(x, t), \quad (43)$$

en el caso de desarrollarla tendríamos una PDF igual a la del FBM, de esta forma, ambos modelos nos darán un MSD igual, aunque tengan una forma de generarse diferente.

Este proceso, al ser tan similar a la difusión normal es característico de ciertos microdominios poco saturados, es decir, microdominios poco definidos y con poca presencia de proteínas o lípidos específicos.



Figura 18: Ejemplo de SBM de 1000 pasos

2.3.3.4.5 Levy Walks

El modelo *levy walks* (LW) comparte similitudes con el modelo CTRW, ya que en su proceso de desplazamiento intercala tantos momentos donde la partícula no realiza ningún salto y otros en los cuales se mueve, solo que en estos casos los saltos serán mucho más pronunciados, desplazando de un punto a otro del medio recorriendo grandes distancias, a una densidad de muestreo baja da la sensación de que se teletransporta (Klafter & Zumofen, 1994).

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

Visto esto, en primer lugar, análogamente al CTRW se definirá una PDF para los tiempos de espera

$$\Psi(t) = 1 - \int_0^t \Psi(\tau) d\tau, \quad (44)$$

y una PDF para los desplazamientos en el espacio, que en este caso será gaussiana como si se tratara de una difusión normal. Siguiendo los pasos análogos al CTRW definiremos una PDF que nos indique la probabilidad de llegar a un punto x en un tiempo t de la forma

$$P(x, t) = \int_{-\infty}^{+\infty} g(y) \int_0^t \Psi(\tau) Q(x - y, t - \tau) dy d\tau + P_0(x) \Psi(t), \quad (45)$$

siendo la $g(y)$ la probabilidad de dar un salto de longitud y , y la función Q , nos indica la cantidad de partículas que dejan el punto x por unidad de tiempo.

La cual empleando las transformaciones de Laplace y Fourier nos dará la expresión general de la PDF

$$P(k, s) = \frac{\Psi(s)P_0(k)}{1 - \Psi(s)g(k)}. \quad (46)$$

Mediante un proceso igual de desarrollo al del CTRW comentado anteriormente, empleando una función de tiempos de espera de Poisson, se llega a un MSD de la forma

$$MSD = \langle x^2(t) \rangle = \int_{-\infty}^{+\infty} x^2 P(x, t) dx = 2Dt, \quad (47)$$

tratándose de un MSD lineal y con coeficiente de difusión constante, por tanto, de muy baja ergodicidad.

Aunque los LW se usen comúnmente para describir la forma de desplazamiento de muchos organismos, tanto unicelulares como pluricelulares, desde amebas en soluciones hasta ballenas en el océano, estos están también relacionados con el mARN, concretamente con el mARN unido a otras proteínas, el llamado mRNP, que se une también a los lípidos de la membrana celular formando pequeñas áreas especializadas de la membrana.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES



Figura 19: ejemplo de LW de 1000 pasos

2.3.3.4.6 Resumen

Modelo	Coefficiente de difusión	Tiempos de espera entre saltos	Distancia total recorrida	Distancia entre saltos	PDF espacial
CTRW	Constante	Alto	Poca	Poca	Varia
FBM	Constante	Muy alto	Media	Media	Gaussiana
ATTM	Varia mucho	Medio	Poca	Poca	Varia
SBM	Varia poco	Bajo	Media	Poca	Gaussiana
LW	Constante	Medio	Mucha	Muchísima	Varia

Tabla 1: Resumen de las características generales de los modelos

2.4 Redes Neuronales Artificiales

En este apartado explicaremos los métodos empleados para trabajar con los datasets obtenidos anteriormente. Con estos métodos trataremos de realizar un clasificador basado en *Deep learning* empleando redes neuronales artificiales.

Hoy en día, el Deep learning representa el estado del arte en prácticamente cualquier proceso relacionado con la minería de datos y la clasificación de imágenes. Dividiendo principalmente esta gran familia en tres grupos diferenciados: Redes densas, redes convolucionales y redes recurrentes.

En este estudio nos centraremos en las dos primeras redes comentadas anteriormente, las redes densas y redes convolucionales (las cuales explicaremos con más detalle en apartados posteriores). Para el análisis de imágenes se ha demostrado que una de las mejores arquitecturas se trata de la unión de una secuencia Autoencoder convolucional (capas convolucionales y capas de max-pooling) concatenado con una red densa de varias capas (Figura 20).

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

2.4.1 Definición redes neuronales convolucionales para clasificación

Llamaremos red neuronal convolucional para clasificación a la unión de red convolucional y densa que se ha nombrado anteriormente (Figura 20)

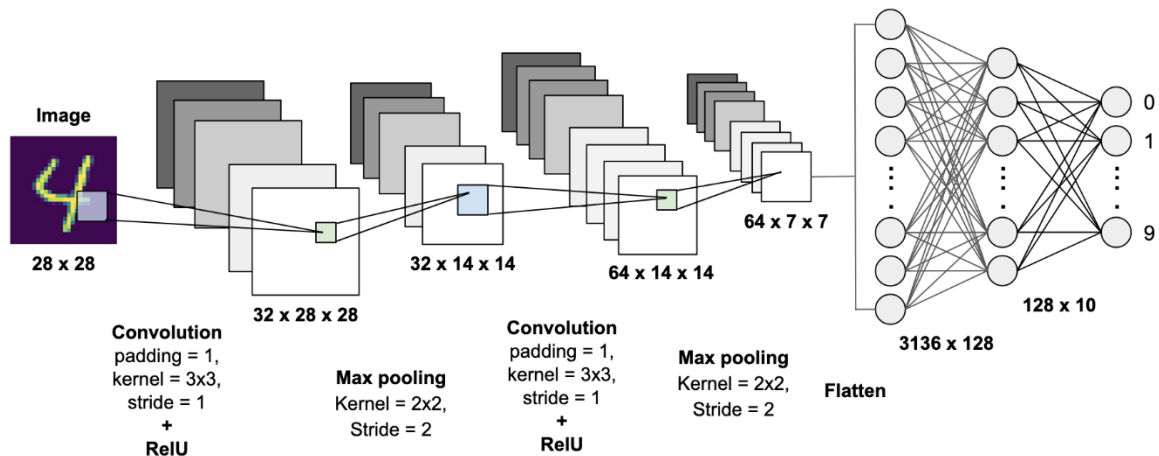


Figura 20: Red convolucional concatenada con una red densa con salida de 9 clases (clasificación de números del 0 al 9, teniendo en este caso como input el número 4) (Shreyak, 2020).

Las redes neuronales convolucionales para clasificación son redes neuronales artificiales de Deep learning basadas en procesos de entrenamiento supervisados. Es decir, estas redes son capaces de generar modelos que distingan entre diferentes clases gracias a un aprendizaje previo en el cual le suministramos pares de muestra y etiqueta. Estas son capaces de extraer características de datos estructurados mediante el aprendizaje de una serie de parámetros o coeficientes (Bagnato, 2018; Garcia Ferrando, Garcia-Gomez & Monserrat, 2017).

Los pares muestra-etiqueta son parte del llamado dataset, es decir, parte del conjunto total de datos que disponemos para el estudio. Los que suministramos durante el entrenamiento son los datos de train, posteriormente para comprobar la eficacia de nuestra red, se suministrarán datos de test, estos datos serán también pares muestra-etiqueta los cuales no habrán sido empleados durante el entrenamiento, lo cual servirá para evaluar con datos independientes el modelo obtenido.

Los pares muestra-etiqueta son conjuntos de elementos que empleamos para entrenar este tipo de redes. Las muestras son los elementos que deseamos que la red aprenda a diferenciar entre sí, es decir, si queremos diseñar una red capaz de distinguir entre perros y gatos a partir de fotos de ambos animales, las muestras serán las propias fotos de perros y gatos. Las etiquetas se tratan del identificador de clase de cada, es decir, lo que nos indica a que clase pertenece cada muestra, siguiendo el ejemplo anterior en este caso las etiquetas serían o perro o gato.

Estas redes cuentan con una arquitectura compleja dividida en capas, estas capas a su vez están divididas en neuronas, siendo la neurona la mínima unidad funcional de este tipo de redes. Las

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

neuronas de cada capa están conectadas con las neuronas de previas y sucesivas capas (el nivel de conexión dependerá del tipo de función que tenga cada una).

2.4.1.1 Arquitectura de la red

Este proceso se divide principalmente en dos fases consecutivas que se repiten en este orden una y otra vez durante el proceso de entrenamiento, se divide principalmente en dos grandes etapas, la etapa convolucional y la densa

2.4.1.1.1 Red Convolucional

La fase convolucional cuenta con neuronas especializadas en este proceso. Tiene tantas capas como compleja deseemos que sea la red y cada capa cuenta con tantas neuronas como casillas tengan las matrices con las que trabajemos. Es decir, si trabajamos con imágenes de $200 \times 200 \times 3$, el total de casillas o píxeles que tendremos será de $200 \times 200 \times 3 = 120K$, por tanto, nuestra primera capa contará con 120K neuronas.

En cada capa se realizarán tres procesos consecutivamente relacionados con la extracción de características, los cuales estarán conectados entre sí.

En primer lugar, se realiza el filtrado. En este paso se emplean kernels o filtros, estos son matrices comúnmente de 3×3 dimensiones y con un solo nivel de profundidad.

	0	1	0	
	0	0	0	
	0	0	0	

Figura 21: Kernel 3×3 con 0-padding (Gimp.org, 2016).

Se realiza el proceso de convolución (Figura 22) entre las matrices de entrada y los kernel, donde el kernel se ira desplazando pixel a pixel realizando la convolución. Esta operación se basa en la suma aritmética de los productos pixel a pixel entre el kernel y la región sobre la que esta superpuesto el kernel de la matriz. En el caso de quedar parte del kernel fuera de la imagen se realizará un zero-padding, es decir, añadir celdas de valor 0 en las zonas donde el kernel quede fuera.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

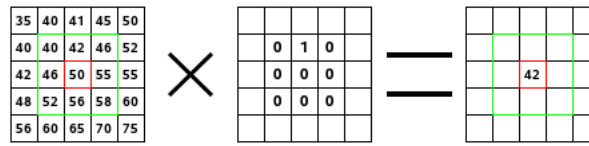


Figura 22: Convolucion del kernel con la casilla central de una matriz (Gimp.org, 2016).

En el caso de tener una primera capa de 32 filtros, es decir, 32 kernels, al pasar la matriz de $200 \times 200 \times 3$ quedaría como resultado una nueva matriz de $200 \times 200 \times 32$, siendo cada nivel de profundidad el resultado de una convolución.

Nota: El número de niveles de profundidad de la matriz empleada no varía el número de niveles de profundidad de la operación de convolución, ya que el paso de un kernel por una matriz siempre producirá una matriz de solo un nivel de profundidad, es decir, una matriz de $200 \times 200 \times 3$ al pasar por un kernel pasara a ser de $200 \times 200 \times 1$.

En segundo lugar, se realizará una fase de activación, es decir, a cada casilla de la matriz obtenida del paso anterior se le aplica una función que tendrá como valor de entrada su valor. Estas funciones se denominan funciones de activación.

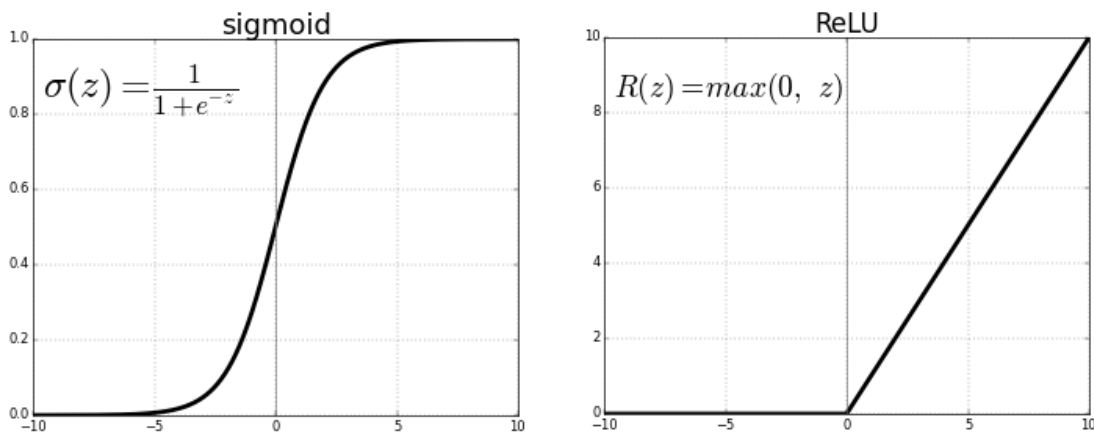


Figura 23: Función de activación sigmoide (izquierda) y ReLu (derecha) (Sharma, 2017).

Las más comúnmente empleadas son las ReLu y la sigmoide (Figura 23). Además, el resultado de este proceso no varía las dimensiones de la matriz, la matriz continuara siendo de las dimensiones $200 \times 200 \times 32$.

$$\sigma(-3) = \frac{1}{1 + e^{-(-3)}} = 0.047 \text{ (Sigmoide)} \qquad R(-3) = \max(0, -3) = 0 \text{ (ReLu)}$$

Por último, realizaremos una etapa de subsampling (muestreo), en la cual se busca reducir la dimensionalidad de las matrices. Para este proceso comúnmente se divide la matriz en sectores del mismo tamaño y se le aplica una función a cada uno, como la media o el valor máximo (Figura 24).

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

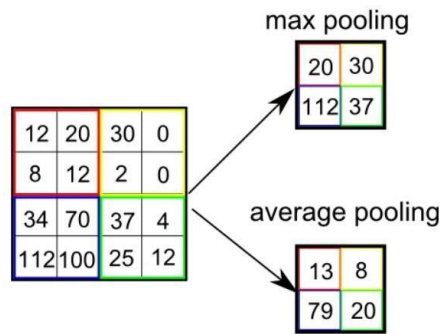


Figura 24: Max pooling y mean-pooling (average pooling) con sectores de 2x2 (Rahman, 2019).

Si por ejemplo empleamos regiones de 2x2 píxeles para realizar el subsampling, en el ejemplo que estamos realizando, al tratarse de regiones de 200x200, de esta forma la reducción se realizara de la forma $200/2 = 100$. Obtenemos como resultado final una matriz de 100x100x32.

Este proceso se realizará tantas veces como capas haya en esta fase. Si por ejemplo contamos con 2 capas más, de 64 y 128 kernels respectivamente, y subsamplings de 2 y 2. En la segunda capa obtendremos una matriz de 50x50x64 y en la segunda 25x25x128.

2.4.1.1.2 Red densa

La fase densa, está compuesta por capas formadas por neuronas densamente conectadas entre sí, es decir, cada neurona de una capa está unida directamente con todas las neuronas de su capa inmediatamente posterior y anterior, siendo la capa anterior la que proporciona los inputs y las posteriores las que reciben los outputs de estas como inputs (Figura 25).

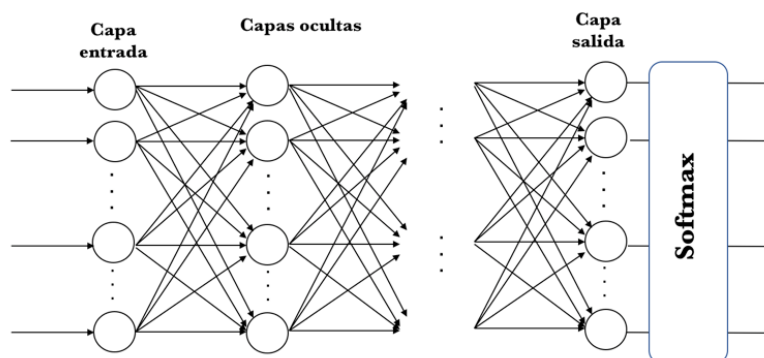


Figura 25: Capas densamente conectadas (Torres, 2018).

En esta fase y la anterior se realiza el aprendizaje, mediante el algoritmo backpropagation o propagación retrospectiva del error. Antes de explicar este algoritmo es necesario explicar las unidades neuronales mínimas que componen estas capas, los perceptrones.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Los perceptrones actúan realizando operaciones matemáticas análogamente a una neurona, siendo sus inputs las dendritas, la zona de cálculo el soma y el output el axón, además de contar con un umbral bias.

Esta unidad actúa como una función multivariable, generando un output a partir de varios inputs. En primer lugar, recibe los inputs y le asigna un peso a cada uno, es decir, realiza el producto entre el valor del input y el peso (w). Posteriormente se suman todos los valores obtenidos y se le resta el valor del umbral bias. Por último, se aplica a este nuevo valor una función de activación, por ejemplo, la función ReLu para obtener el output final.

$$A = F.act\left(\sum_{i=0}^n x_i w_i + b\right)$$

El umbral *bias* es muy importante ya que mediante la suma de este se puede desplazar la función de activación de izquierda a derecha, o mediante su producto se puede inclinar esta función, consiguiendo así un mejor ajuste de los datos.

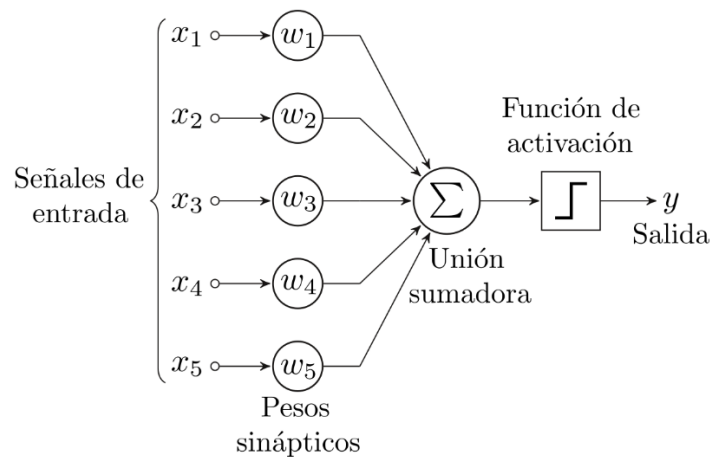


Figura 26: Perceptrón con 5 entradas (Torres, 2018).

Todas las neuronas de las capas densas serán perceptrones, solo que cada perceptrón tendrá tantas salidas como neuronas haya en la siguiente capa. Las capas tendrán una reducción sucesiva en su número de neuronas según avancemos en estas, hasta llegar a la capa final o softmax, la cual tendrá tantas neuronas como clases tengamos que clasificar, en el ejemplo de perros y gatos tendrá 2. Los pesos de esta última capa representaran los porcentajes de pertenencia a cada clase, siendo por ejemplo la clase 1 perros y la 2 gatos, si diera como resultado $[0.8, 0.2]$ significaría que la muestra dada tiene un 80% de ser perro y 20% de ser gato, por tanto, la clasificaremos como perro.

El cálculo de los valores finales que generará nuestra red en cada nodo de salida será el resultado de la propagación de las muestras por cada uno de los nodos. Siendo la expresión general para cada nodo

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

$$A_j^{(i+1)} = F. act\left(\sum_{k=0}^n w_k^{(i+1)} x_k^i + b^{(i+1)}\right),$$

siendo A el output del perceptrón, el superíndice la capa en la que se encuentra y el subíndice la neurona.

Haciendo este cálculo recursivo hallamos el valor final predicho por la red, también llamado valor hipótesis, el cual tendrá un valor para cada neurona del softmax.

2.4.1.1.3 Sistema de aprendizaje

El sistema de aprendizaje se desarrolla con el empleo de los algoritmos forward propagation y back propagation, afectando tanto a la fase convolucional como densa. Este sistema se basa en la corrección sucesiva de los pesos de los perceptrones empleando la reducción de una función de coste. Esta función cuenta con la expresión general

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m Cost(h_{\theta}(x^{(i)}), y^{(i)}),$$

siendo m el número total de muestras empleadas en el entrenamiento, θ los pesos de cada nodo y $Cost()$ el coste de la función. Esta última función es la que marcara la diferencia entre el valor obtenido en el nodo un nodo concreto x y su valor real y , se suele emplear el error como diferencia entre ambos, es decir, la resta de ambos valores tal que

$$e_j(n) = d_j(n) - y_j(n).$$

Los valores que pueden tomar tanto el valor deseado como el valor calculado están en el intervalo $[0,1]$. Los valores deseados en la última capa softmax tendrán valores 0 o 1, ya que se trata de la etiqueta que ha fijado la red. En nuestro caso, que contamos con 5 neuronas de salida, ya que tenemos 5 clases a clasificar, si por ejemplo mandamos un par muestra-etiqueta de la clase CTRW, que corresponde a la primera clase, la primera neurona de la capa softmax arrojará un valor 1 y las cuatro restantes un valor 0. En el caso de los valores predichos por la red, estos oscilarán en el rango $[0,1]$ en la capa softmax, siendo el objetivo de la red hacer que estos se asemejen lo más posible a los valores deseados discretos.

$$0 \leq h_{\theta}(x) \leq 1$$

El objetivo por tanto es optimizar la función de coste, de forma que para encontrar la solución ideal la red tendrá que minimizar al máximo el valor de la función de coste J.

En muchos casos la función J, se trata de una superficie convexa llena de irregularidades en el espacio, contando con muchos mínimos locales. De esta forma, no se puede garantizar que la función de coste llegue al mínimo global, pudiendo quedarse atascada en algún mínimo local. Para evitar este problema la red toma logaritmos del coste de la forma:

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

$$cost(h_{\theta}(x), y) = \begin{cases} -\log(h_{\theta}(x)) & \text{si } y = 1 \\ -\log(1 - h_{\theta}(x)) & \text{si } y = 0 \end{cases}$$

Mediante esta función, la cual depende del valor deseado que fijamos mediante las etiquetas de las muestras, evitamos que se converja en mínimo locales. Ambas funciones presentaran la siguiente forma general:

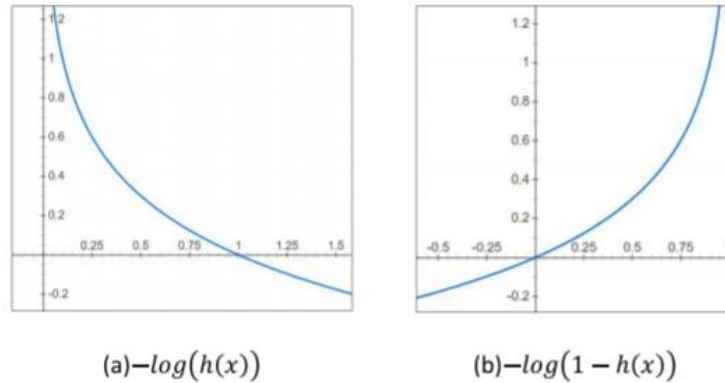


Figura 27: Funciones logarítmicas graficadas

En el primer caso, cuando el valor se aproxime a 1, esta función tenderá a tomar valores cercanos a 0 y por tanto tendrá un coste muy bajo, en el caso que se aproximen a 0, tomará valores que tenderán al infinito y por tanto tendrán un coste muy alto, el cual indicara que se nos estamos alejando del mínimo global de la superficie.

En el segundo caso, actúa de forma contraria, tomando valores cercanos a 0 cuando el valor predicho tome valores cercanos a 0 y viceversa con los valores cercanos a 1.

Dado que el valor y es siempre 0 o 1 de forma discreta, se puede expresar la función de la forma

$$Cost(h_{\theta}(x), y) = -y \log(h_{\theta}(x)) - (1 - y) \log(1 - h_{\theta}(x)),$$

si sustituimos en la ecuación general de la función de coste obtenemos

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[y^{(i)} \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})) \right] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2,$$

en el caso de disponer de más de un nodo de salida en la capa final softmax, como es nuestro caso, debido a que se quiera hacer una clasificación de muchas clases, la expresión quedaría de la forma

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[y_k^{(i)} \log\left(\left(h_{\theta}(x^{(i)})\right)_k\right) + (1 - y_k^{(i)}) \log\left(1 - \left(h_{\theta}(x^{(i)})\right)_k\right) \right] + \frac{\lambda}{2m} \sum_{l=1}^{L-1} \sum_{i=1}^{S_{l+1}} \left(\theta_{j,i}^{(l)}\right)^2,$$

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

siendo K el número de nodo de esta capa.

Una vez definida esta función de coste, la cual tendrá un coste basado en el error o diferencia de valores en nuestro estudio, podemos definir los dos pasos consecutivos que realiza esta red, el forward propagation y back propagation.

Durante el forward propagation se realiza una pasada sobre las capas en dirección a la última capa o softmax, la cual tiene como objetivo generar un valor de salida para cada muestra, en nuestro caso expresado como un vector de 5 posiciones de porcentajes relativos en el rango $[0,1]$.

Una vez obtenidos todos los valores predichos, se calcula el coste general mediante la ecuación de coste y se almacena para el siguiente apartado.

En el proceso del back Propagation se busca minimizar el valor de la función de coste obtenida en el apartado anterior, esto se consigue reajustando los pesos de cada capa recursivamente. Para esto se emplea el gradiente de la función de coste, es decir, su derivada parcial respecto a la variable pesos. Una vez obtenida esta nueva función gradiente se emplea el algoritmo de descenso de gradiente, el cual recalcula el nuevo peso mediante la diferencia de su valor con respecto al valor del gradiente en ese peso multiplicado por la tasa de aprendizaje o learning rate.

Algoritmo de descenso de gradiente:

1. *Repedir*{

$$2. \theta_j := \theta_j - \alpha^{11} \frac{\partial}{\partial \theta_j} J(\theta)}$$

$$3. \textit{detente cuando: } \frac{\partial}{\partial \theta_j} J(\theta) \rightarrow 0$$

Esta reasignación de pesos se realizará hacia atrás, es decir, empezando por la capa softmax y acabando en la primera capa densa, pasando de forma ordenada por todas las capas ocultas. Esta propagación hacia atrás favorece en el sucesivo paso de forward propagation ver como contribuye cada peso al error general. Los pesos que más contribuyen al error tendrán valores de derivación mayores y por tanto serán los que más modifiquen su valor.

2.4.1.2 Diseño de la red

La arquitectura de la red que diseñemos debe ser la óptima para la resolución de nuestro problema. Para ello debemos de seleccionar como desarrolladores tanto el número de capas convolucionales y densas que deseamos, además del número de neuronas de cada capa, el tamaño de los filtros kernel de convolución, la ratio de aprendizaje, los batches y las épocas; lo cual llamaremos hiperparametros. También tendremos en cuenta otros dos aspectos muy importantes tales como el batch normalization y el sobreentrenamiento o overfitting.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

2.4.1.2.1 Hiperparametros

El diseño de este tipo de redes contara con una serie de hiperparametros que deberemos introducir nosotros con tal de optimizar su funcionamiento, los cuales se basaran en el conocimiento del problema por parte de los investigadores. Principalmente encontramos los siguientes:

- **Batch:** Este parámetro marca el número de muestras del total del dataset que se introducen en cada época, es decir, en cada proceso de actualización de los parámetros internos del modelo, de los pesos más exactamente.
- **Epochs:** Se trata del número de veces que se ejecutara el algoritmo de aprendizaje, es decir, las sucesivas fases de forward propagation y back propagation, además de al emplear batches de muestras diferentes del total de muestras también se realizara la fase de extracción de características convolucional.

Debe de haber un numero óptimo de épocas tanto para no quedarse cortos, de forma que no se conseguiría reducir el error todo lo deseado, o pasarse y producir un efecto de sobrenetrenamiento (overfitting)

- **Learning rate:** Este parámetro aparece en el algoritmo de descenso de gradiente, generando un producto entre este y el gradiente de la función de coste. Este cuantifica como de rápido va a ser el proceso de descenso y búsqueda del mínimo global, es decir, como de largos serán los pasos de avance que se emplearán para recorrer la superficie de la función.

La elección de un learning rate optimo es crucial, ya que si elegimos uno demasiado pequeño este puede hacer que la búsqueda del mínimo global se atasque en un mínimo local, en el caso de elegir uno demasiado alto puede hacer que durante el proceso de búsqueda los saltos sean demasiado grandes y no sean capaces de llegar al mínimo, pasándolo de largo y oscilando continuamente alrededor de este.

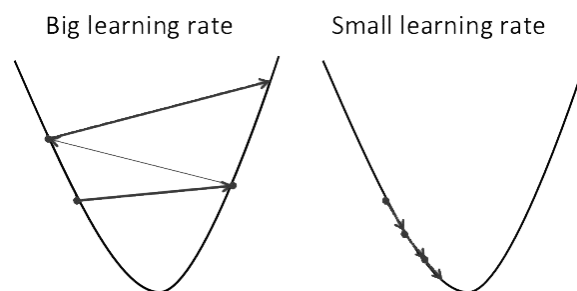


Figura 28: A la derecha learning rate demasiado alto y la izquierda learning rate demasiado bajo (educative.io, 2017).

2.4.1.2.2 Batch Normalization

Lo añadimos tanto en las capas convolucionales como en las densas. Durante los procesos de entrenamiento los parámetros ocultos que se encargan de ajustar los pesos pueden moverse gradualmente. Con el batch normalization normalizamos los inputs de cada capa ajustando y escalando las activaciones. Centrando su valor en medias de 0 y desviaciones de ± 1 . Estos parámetros no serán

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

fijos, podrán modificarse ligeramente si la red considera que esto será beneficioso para el entrenamiento.

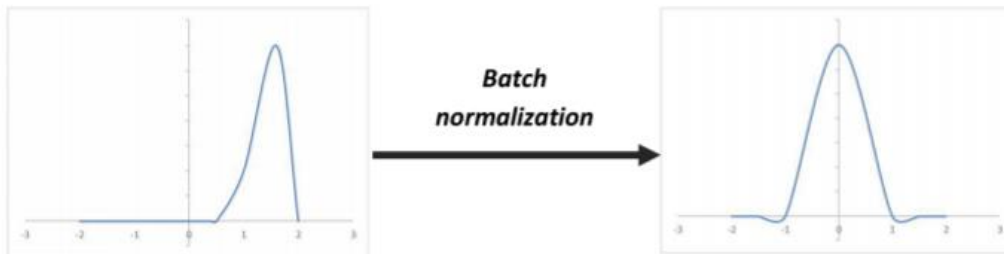


Figura 29: Proceso de batch normalization sobre distribución de Poisson (Ioffe & Szegedy, 2015).

2.4.1.2.3 Sobreentrenamiento

El overfitting o sobreajuste se produce cuando la complejidad de la red no permite identificar características generalizadas, sino que solo ajusta sus pesos para describir el conjunto de train en su entrenamiento. Esto quiere decir que al procesar sobre esta red cualquier otro conjunto de datos no los clasificará correctamente ya que no habrá extraído las características generales de cada una de las clases.

Es muy importante controlar este factor, esto se consigue definiendo una arquitectura de red adecuada para nuestro problema. Aunque se puede tener a pensar que una arquitectura más compleja siempre va a ser más beneficiosa ya que extraerá un mayor número de características que después se podrán clasificar, esto es incorrecto, ya que muchas de estas características pueden no ser generales de los modelos, sino de las muestras concretas que le estamos suministrando durante el entrenamiento.

Un elemento que podemos emplear para evitar este sobreentrenamiento es el dropout. En este proceso se especifica una probabilidad, por ejemplo, $p=0.5$, esto querrá decir que en cada época de entrenamiento cada neurona de cada capa densa tendrá una probabilidad de desconectarse (no recibir ningún input ni generar ningún output) del 50%.

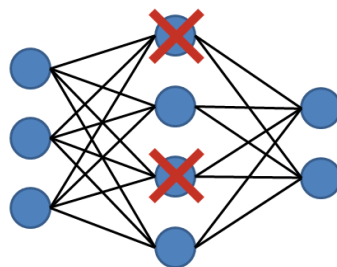


Figura 30: Dropout $p=0.5$ en la capa profunda (Dubbura, 2018).

3. Materiales

Para la realización de este trabajo se han empleado materiales no tangibles, es decir, algoritmos y sistemas informáticos. Debido a su carácter informático no se ha hecho uso de ningún laboratorio ni material experimental.

3.1 Código generador de trayectorias

Disponemos de código para generar random walks de los 5 modelos estadísticos explicados anteriormente. Estos códigos son suministrados por el ANDI Challenge y han sido mejorados por nosotros y otros participantes de la challenge.

En primer lugar, disponemos de un código con los algoritmos necesarios para generar cada uno de los cinco modelos explicados anteriormente en 2D, en la subsección de modelos anómalos de modelos difusivos, empleando las ecuaciones de cada modelo para su generación.

Por otro lado, disponemos de un código main. Este código es el encargado de gestionar la generación de estos modelos mediante los algoritmos. Este archivo nos permite generar datasets de cada modelo fijándoles una serie de parámetros, dependiendo de cómo deseemos que sean las trayectorias que obtendremos. Estos parámetros son los siguientes:

- **Longitud de la trayectoria:** Numero de pasos que queremos que hagan las partículas durante la generación del RW.
- **Coefficiente anómalo (α):** parámetro que regirá el MSD de las trayectorias.
- **Numero de trayectorias:** Numero de trayectorias que tendrá nuestro dataset generado.
- **Modelo:** Modelo del que deseamos generar las trayectorias del dataset.
- **Nivel de ruido:** A modo de acercarse lo máximo posible a la realidad es conveniente generar un ruido gaussiano, que es el que recibiríamos de estar empleando una técnica de captación SPT como se ha explicado anteriormente. Por tanto, definiremos los parámetros de la gaussiana media y desviación estándar, que serán aplicadas a cada punto de la trayectoria.

3.2 Servidor de cómputo

Servidor externo de la unidad de Bioinformática y Bioestadística del instituto de investigación príncipe Felipe.

- Frecuencia de CPU de 2'1 GHz, es decir, capaz de realizar 2100 millones de ciclos por segundo.
- GPU Nvidia Tesla tesla v100-PCIE-32GB.
- Sistema de encolamiento de trabajos.
- 44 nodos con 660 CPUs

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

- Memoria RAM de 11 Terabytes y 1 petabyte de capacidad en disco duro.

Nos conectamos al de forma remota mediante VPN en sistema operativo Ubuntu (ejecutado sobre máquina virtual de Linux).

3.3 Lenguajes y tecnologías

Para el desarrollo del código se ha implementado el mismo lenguaje que el de los códigos suministrados por el *challenge*, Python, en su versión 3.4. Como entorno de programación se ha empleado el entorno Spyder de Anaconda.

Para el desarrollo de las redes convolucionales se ha empleado la librería libre y gratuita TensorFlow, que cuenta con todos los métodos necesarios para el desarrollo de estas redes paso a paso, empleado la versión 2.0.

Para la codificación de algoritmos en las GPUs del servidor de cómputo empleamos la plataforma CUDA, empleando wrappers para su adaptación de lenguaje C a Python.

4. Métodos

Los métodos empleados, debido a que se trata de un proceso secuencial, seguirán un orden a la hora de implementarlos durante la fase de experimentación. El orden será el mismo que el de los apartados mostrados.

4.1 Minería de datos sobre trayectorias

Como en cualquier estudio basado en el análisis de señales o datos, será necesario realizar un preprocesado previo de los datos. De esta simplificamos la información extrayendo de los datos crudos las características más favorables para nuestro estudio, en nuestro caso aquellas que ayudaran más a clasificar las trayectorias entre modelos. De esta forma, presentamos los datos de una forma que la red sea capaz de tratarlos más fácilmente.

4.1.1 Generación trayectorias

Para facilitar el posterior almacenamiento de los datos se han generado los datos de cada modelo por separado, es decir, 5 datasets distintos uno para cada modelo: CTRW, ATTM, SBM, LW y FBM.

Debido a que en la realidad la longitud en la que se nos presentara cada trayectoria variara, es decir, a una misma densidad de muestreo, aunque empleemos los mismos marcadores para todas las moléculas, estas no tendrán la misma vida, algunas se extinguirán antes que otras y por tanto dispondremos de trayectorias con diferentes longitudes, que podrán distar mucho unas de otras.

No es conveniente recortar las trayectorias a una longitud constante fijada por nosotros, debido a dos problemas principalmente. En primer lugar, nunca sabremos cual será la longitud real de las trayectorias que recibiremos al realizar SPT sobre células en un futuro, de forma que podríamos estar empleando trayectorias fijas demasiado largas y tener que interpolarlas, lo cual induciría mucho nivel de ruido. En segundo lugar, dependiendo del coeficiente α , la respuesta subdifusiva o superdifusiva característica de cada modelo puede manifestarse con mayor o menor retardo, siendo imposible saber en qué lugar sería el óptimo para recortar las trayectorias, haciendo que en muchos casos si se recorta en puntos iniciales de esta, presente un comportamiento puramente difusivo normal, y en caso de recortarse en tramos finales, que esté presente un comportamiento puramente logarítmico o exponencial, el cual no representa ningún tipo de difusión estudiada.

Visto esto, nuestro método debe ser capaz de discriminar entre trayectorias de diferentes longitudes mezcladas entre sí. Para ello se ha desarrollado un nuevo código capaz de generar estos nuevos datasets mezclados entre sí.

Esta nueva función llamara a la función main del código suministrado por el challenge, ejecutándola tantas veces como sea necesario para generar estos nuevos datasets. Los parámetros que le deberemos de suministrar serán los siguientes:

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

- **Intervalo de longitudes:** longitud de inicio, longitud final y saltos entre longitudes. Es decir, en el caso de querer longitudes de 100, 200, 300 y 400 puntos, los parámetros tendrán un valor de 100, 400 y 100 respectivamente. Empezara en 100 y saltara de 100 en 100 hasta llegar a 400.
- **Intervalo de α :** alfa de inicio, alfa final y saltos entre α . Análogo al parámetro anterior, pero con el coeficiente anómalo.
- **Repetición de α (N):** Número de veces que se repite cada α . Es decir, si disponemos de 3 alfas: 0'5, 0'6 y 0'7 y una N=3, dispondremos en total de 3 trayectorias de 0'5, 0'6 y 0'7, un total de 9 trayectorias.
- **Modelo:** Modelo del que deseamos generar las trayectorias.
- **Niveles de ruido:** media y desviación estándar de la gaussiana.

Esta función la hemos nombrado como `create_model_dataset ()`.

Este algoritmo, junto a las posteriores funciones se ejecutará dos veces, una para la creación del dataset de train y otra para el de validación, mediante dos funciones por separado que se encargaran de generar dos datasets independientes.

4.1.2 Generación matrices

Debido a la gran adaptación de las redes CNN en el uso de imágenes se ha decidido tratar estas trayectorias almacenadas en los datasets como matrices bidimensionales.

Este tratamiento tiene un problema principal, se trata de la estocástica de las trayectorias. Estas trayectorias como bien indica su nombre Random Walks, son trayectorias aleatorias, presentaran por tanto tantas distribuciones en el espacio como trayectorias haya . Esto, como en cualquier imagen que no podamos predecir su distribución espacial o registrarla, dificulta en gran medida el problema.

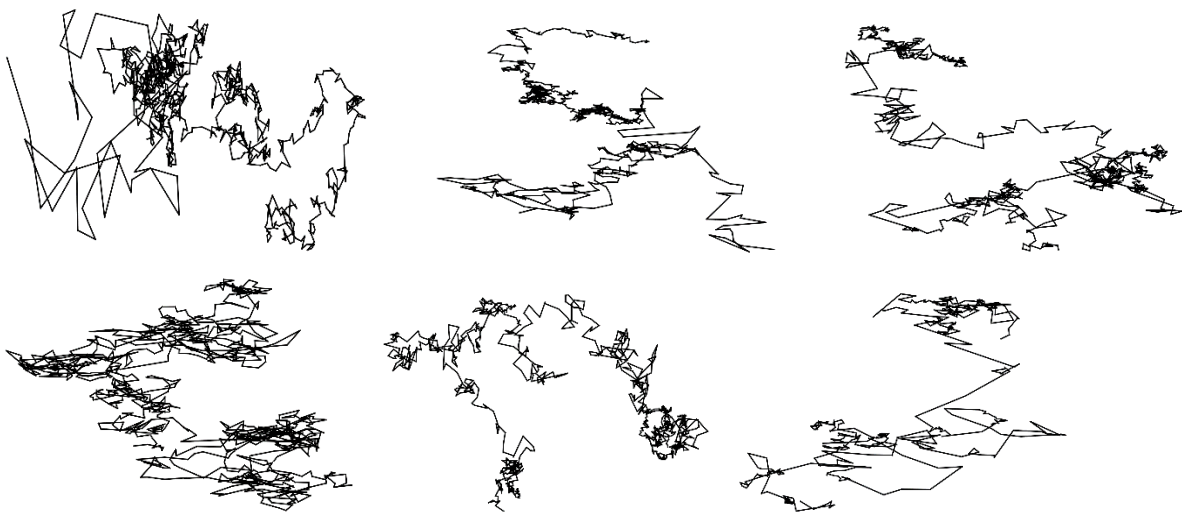


Figura 31: Seis trayectorias diferentes generadas con el modelo CTRW, tienen la misma longitud pero diferente forma y distribución en el espacio

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Debido a este inconveniente la información que se pase a la CNN deberá de estar codificada de forma que se le facilite encontrar fácilmente descriptores que no estén basados principalmente en su distribución espacial, sino más bien en sus características intrínsecas, en nuestro caso, las características propias de los modelos generativos de cada trayectoria. Las matrices con las que trabajemos tendrán por tanto que añadir información extra además de la propia información espacial.

Para la codificación espacial se han generado matrices cuadradas de dimensiones fijas, en nuestro caso de 200×200 pixeles, debido a los límites de almacenamiento que disponemos, es un tamaño que nos permite trabajar con grandes volúmenes de datos, además de que esta cantidad de pixeles nos otorga una definición bastante buena de las trayectorias.

Para la generación de las matrices realizaremos los siguientes pasos:

-Tomamos una trayectoria del dataset a emplear, es decir, extraemos de una posición de la lista una matriz bidimensional y de esta extraemos una fila completa con las coordenadas x e y .

-Calculamos los puntos más extremos de cada trayectoria, el más extremos por la derecha, por la izquierda, arriba y bajo. Estos puntos definirán los límites de nuestra imagen, es decir, estas coordenadas serán las que toquen cada uno de los 4 bordes de la matriz de 200×200. Llamaremos máximo y mínimo a lo que correspondería a las coordenadas negativas y positivas de la X e Y en el plano cartesiano. Teniendo X_{max} , X_{min} , Y_{max} e Y_{min} .

-Ahora toca iterar con todos los puntos de cada trayectoria, hasta que esta se acabe.

Tomamos el punto en el que nos hallemos en ese momento, su coordenada x e y (si acabamos de empezar tomamos el primer punto)

Calculamos la posición x e y nuevas de las coordenadas en la matriz restándole a cada valor el valor de la x e y mínimas y realizando el producto de esta diferencia con el valor de las dimensiones de la matriz, el llamado factor de escalada, en nuestro caso 200, posteriormente redondeamos este valor y los dividimos entre la diferencia entre el valor máximo y mínimo para obtener la casilla que ocupara está en la matriz final. El algoritmo sería el siguiente:

```
 $x\_range = max\_x - min\_x$   
 $y\_range = max\_y - min\_y$   
 $if\ x\_range == 0:$   
     $x\_range = 2 * min\_x$   
     $min\_x = 0$   
 $if\ y\_range == 0:$   
     $y\_range = 2 * min\_y$   
     $min\_y = 0$   
 $x\_start = int(round((scaling\_factor * (x0 - min\_x)) / x\_range))$   
 $y\_start = int(round((scaling\_factor * (y0 - min\_y)) / y\_range))$ 
```

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Aplicamos lo mismo para el punto inmediatamente siguiente, es decir, el (x_{i+1}, y_{i+1}) , obteniendo así también su posición en las casillas de la matriz.

Unimos ambos puntos mediante una línea recta empleando el método line de Python, el cual dibuja una línea lo más recta posible entre dos casillas de una matriz.

Repetimos el proceso hasta completar toda la trayectoria, es decir, para trayectorias de L de longitud, L-1 veces.

Realizamos este proceso para generar el esqueleto de las matrices, es decir, las zonas de la imagen que tendrán información, de esta forma cargaremos la información adicional sobre este esqueleto de trayectoria. Las zonas por las que no haya pasado la trayectoria tendrán un valor de 0 en la matriz.

4.1.3 Definición variables

Añadiremos información conjuntamente a las trayectorias generadas en el apartado anterior, esta información debe ser muy descriptiva y ser capaz además de diferenciar entre modelos. Esta información además debe estar lo más comprimida posible, es conveniente emplear un pequeño número de matrices para describir cada trayectoria, además de que deben estar ambas lo menos relacionadas entre sí, estos se debe a que en el caso de que varias matrices expresen una información similar entre ambas, la red puede extraer de ellas los mismos descriptores y por tanto tener parámetros redundantes, teniendo neuronas por duplicado y por tanto que muestren discrepancia entre sí para la asignación de los pesos. Por ejemplo, si le suministramos a una imagen con dos niveles de profundidad, en el que el primero se exprese el gradiente de cada pixel respecto a sus pixeles de alrededor y en otra las direcciones de máxima variación de la imagen, posiblemente mediante cualquiera de las dos capas la red será capaz de inferir el parámetro dirección máxima global de variación.

4.1.3.1 Orden en el tiempo

Generamos matrices con una trayectoria de esqueleto, empleando números ordenados con la misma distancia entre si para expresar el paso del tiempo, lo que correspondería a los diferentes niveles de gris en una imagen binaria.

El algoritmo actuará a la vez que se van generando las imágenes de trayectorias, añadiendo un valor al pixel y la línea de unión entre pixeles en cada momento que es añadido a la matriz. El valor de este pixel será igual al valor de la posición que ocupa este paso en el total de las trayectorias, es decir, si la trayectoria tiene 1000 pasos, y el paso actual es el número 47, la casilla que ocupe este paso en la matriz tendrá un valor de 47, además de todas las casillas de la línea recta (distancia euclídea mínima) que lo una con el pixel 48.

Debido a que trabajamos con trayectorias de diferentes longitudes, y podría inducir a error emplear diferentes rangos de valores, debido a que la red podría intentar diferenciar modelos en base a los

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

diferentes rangos de valores de cada trayectoria, y en base a sus características intrínsecas. Para solventar este problema normalizamos las trayectorias respecto al valor máximo de la longitud, es decir, dividiremos cada valor entre el número total de pasos, de la siguiente forma:

$$Imagen[x,y] = \frac{t + 1}{T}$$

Siendo T el número total de pasos y t el instante en el que se encuentra la partícula en la posición (x, y)

Los resultados por tanto si los graficáramos como una imagen binaria de grises quedaría de la siguiente forma:



Figura 32: Trayectoria del modelo ATTM codificada en tiempo

4.1.3.2 Distancia entre saltos

Análogamente a el apartado anterior, la variación de las trayectorias en su forma de desplazarse por el espacio es muy descriptivas en cuanto a modelos, especialmente la longitud que hay entre saltos sucesivos, los cuales la red puede comprobar cuál es el orden de estos mediante la codificación anterior.

Debido a que todas las trayectorias están escaladas, es decir, independientemente de las dimensiones de la región del espacio que estas ocupen, serán expresadas dentro de un marco común, todas ocuparan al completo las matrices de 200×200 , por tanto, será imposible diferenciar cual es la distancia real recorrida entre pasos. Esto es debido a que cada modelo presenta un rango de desplazamiento diferente, algunos como el CTRW se enmarcan dentro de un rango que generalmente no supera las 5 unidades de desplazamiento máximo entre punto de partida y punto más alejado, otros como el LW pueden alcanzar distancias de hasta 300 unidades. En el caso de no escalar estas trayectorias, las más cortas quedarían ocultas como un punto en el centro de la imagen.

Por tanto, consideramos oportuno codificar esta distancia entre pasos, en concreto emplearemos la distancia euclídea. Análogamente a la codificación anterior, se generará alternativamente con las

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

trayectorias. A cada punto se le asignara el valor de la distancia euclídea real que tiene este paso con su anterior, y así sucesivamente, además de que este valor también se les asignara a todas las casillas de la línea que une ambos puntos.

$$Imagen[x_i, y_i] = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}$$

Los resultados no se pueden graficar debido a que la matriz almacena datos decimales.

4.1.3.3 Tiempo máximo de espera

La mayoría de las trayectorias cuentan con pasos en los que las partículas no experimentan ningún tipo de movimiento, es decir, esta se para durante unos instantes en una misma posición, debido a que este suceso es físicamente imposible, además del ruido gaussiano que presentan las máquinas de SPT, nunca encontraremos una partícula parada en una coordenada exacta del espacio, sino dentro de una pequeña región.

Para ello tomamos un radio que definirá la región de confianza, es nuestro caso 3 (1 corresponde a la longitud de un lado de una casilla de la matriz). Es decir, consideraremos que una partícula esta parada si esta no se desplaza fuera de una circunferencia de diámetro 6 (en distancia euclídea).

Recorreremos cada trayectoria punto a punto, definiendo para cada punto este radio y contando la cantidad de instantes que pasan hasta que algún paso sucesivo escapa de este radio. Por ejemplo, si estamos en el instante 1, y la partícula no sale de este radio hasta el instante 4, el valor de tiempo de espera del instante 1 será 4.

Calcularemos este parámetro para todos los instantes y sacaremos el máximo de ellos. Este será el tiempo máximo de espera.

La mayoría de las modelos presentan un valor similar, que se mueve dentro de unos rangos semejantes, en cambio, el modelo CTRW y FBM contienen valores muy superiores a los del resto, estando la media del FBM en valores alrededor de 280 y en el CTRW alrededor de 25, mientras que en los demás modelos no superan valores de 5.

Este parámetro además es muy complicado de inferir por la red de forma natural en un entrenamiento, ya que conlleva mezclar tanto la codificación en el espacio, que denota el orden de los pasos, como la codificación en el tiempo, que denota la distancia entre pasos, además de que la red no está implementada para aplicar parámetros tales como umbrales.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

4.1.3.4 Desviación estándar del coeficiente de difusión aproximado

Como ya se ha explicado en el apartado de teoría, los modelos de difusión brownianos no solo están regidos por la variación en el tiempo y el espacio de su función de posición, sino también por el coeficiente de difusión que indica con que facilidad se podrá mover por el medio (ecuación de Einstein).

La variación del coeficiente de difusión puede resultar muy interesante, ya que nos ayuda a distinguir entre el modelo ATTM y el resto de los modelos. Como se ha explicado en el apartado de teoría, el modelo ATTM actúa de forma muy similar al CTRW, solo que presentando variaciones en su coeficiente de difusión D .

El coeficiente de difusión es muy difícil de calcular, y mucho más si no se dispone información sobre el medio, nosotros solo disponemos del movimiento de la partícula, por tanto, lo aproximaremos mediante el tMSD.

Para calcular este coeficiente se calcula el tMSD de cada trayectoria, nosotros hemos empleado intervalos de tiempo de 1 a 9 instantes, por tanto, dispondremos de 9 puntos en el espacio. Aproximamos estos puntos a una recta mediante una interpolación lineal simple y calculamos la pendiente de esta recta. La pendiente de esta recta será el coeficiente de difusión teóricamente, aunque en estos casos será aproximado ya que nunca formaran una recta perfecta los puntos, sino que presentara desviaciones al final en los casos sub y superdifusivos.

Para apreciar la variación de los coeficientes de difusión a lo largo de las trayectorias hemos decidido cortar las trayectorias en trozos, en nuestro caso 5 trozos de la misma longitud, y calcular el coeficiente de difusión aproximado para cada segmento. Una vez calculado estos 5 coeficientes de difusión se calcula la desviación típica entre los 5 y se almacena este valor.

Como en el modelo ATTM nos daban valores del orden de potencias a la 20, en el resto de los casos valores del orden a la -5. Estos valores son demasiado grandes para poder trabajar con ellos cómodamente, así que les decidimos aplicar una transformación logarítmica empleando el logaritmo en base 10, lo cual nos aproxima los valores con potencias a las -5 a 0 y nos rebaja aquellos con potencias a la +20 a valores cercanos a 40, por tanto, reduciremos el espectro a $[0,45]$ aproximadamente, mucho más cómodo para la red de trabajar.

4.1.3.5 Normalización

En muchos casos será necesario normalizar nuestras trayectorias, dependiendo de cómo se introduzcan estas en los datasets, ya que estas interactuarán con el resto de las variables de los modelos.

Durante el proceso de entrenamiento en todas las capas se aplica un proceso de batch normalization (lo explicaremos posteriormente), que centrará sus valores en una media de 0 y una

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

desviación típica de ± 1 , es decir, la mayoría de los valores estarán en el rango $[-1,1]$ salvo excepciones como outliers. Por tanto, convendrá normalizar nuestros valores a un rango similar, como nuestros valores serán siempre positivos, los llevaremos al rango $[0,1]$.

Para esto realizamos un estudio previo de los valores que pueden adoptar estas variables, de esta forma podremos normalizarlos respecto al valor más alto dentro de unos límites, no respecto al valor más alto absoluto, ya que este podría ser un outlier e introducir ruido durante el normalizado.

Para evitar este problema de outliers empleamos realizamos una previa limpieza de estos, empleamos el método de Tukey's fence, este se basa en la distancia intercuartílica. Calculamos el primer y tercer cuartil (correspondientes a el 25% y 75% de los datos respectivamente) y calculamos por la distribución los siguientes límites (Zhou, Shen, Li, Li, Wang & Wang, 2008):

$$\text{Rango} = (Q1 - k(Q3 - Q1), Q3 + k(Q3 - Q1))$$

Empleando una $k=1.5$, todos aquellos valores fuera de este rango serán considerados outliers y por tanto eliminados de la lista para el estudio.

Una vez hemos limpiado la lista de outliers calculamos el valor máximo para ambos casos, siendo en el caso de la varianza un valor de 31,92 y en el caso del tiempo máximo de espera 350.

Para la varianza tomaremos por tanto un valor de normalizado de 32, ya que los 4 modelos que no son el ATTM tienen valores máximos cercanos a 4. En cambio, el valor de espera máximo no convendrá tomar un valor tan alto como 350, ya que este corresponde al modelo FBM, mientras que el modelo CTRW cuenta con unos valores máximos cercanos a 35, y el resto cercanos a 5. Por tanto, al normalizar entre 350, los valores de CTRW y los modelos SBM, LW y ATTM, podrían quedar demasiado próximos entre sí y perder así la descripción del CTRW, así pues, normalizaremos respecto a 200 para alejar más el CTRW de los demás modelos.

Todo valor que se exceda de este se le asignará un valor de 1, y todos aquellos que estén por debajo se dividirán entre este.

4.1.4 Formación de datasets

Los datasets finales se definirán en función del tamaño de estos, a modo de realizar experimentos con una métrica de precisión lo más realista posible, estos datasets serán del mismo tamaño para los 5 modelos.

Para que todos tengan el mismo número de trayectorias en ellos, definiremos intervalos iguales de alfas, es decir, el mismo número de alfas para cada dataset, aunque cada modelo tenga un rango de alfas diferentes. Para ello emplearemos saltos entre α de 0'1 y los siguientes rangos para cada modelo:

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Modelo	α_{min}	α_{max}
CTRW	0.5	1
ATTM	1.1	1.6
LW	1	1.5
SBM	1	1.5
FBM	0.5	1

Tabla 2: Intervalos de α por modelo

El rango de longitudes será común a todas las trayectorias, se comprimirá entre 200 y 800 con saltos de 100, es decir, tendremos trayectorias de: 200, 300, 400, 500, 600, 700 y 800 en todos los modelos.

El parámetro que definiremos por tanto será el parámetro N (trayectorias de mismo modelo con mismo α), este parámetro dependerá del almacenamiento del que dispongamos, según el experimento que realicemos cada muestra tendrá un peso concreto y por tanto tendremos que hacer datasets con una N más reducida. De normal trabajaremos con datasets de N entre 100 y 200.

Para esta tarea no emplearemos archivos tensor como se suelen emplear comúnmente, sino archivos TFRecord, estos archivos tienen una ventaja respecto a los tensores que es su capacidad de ser cargados en el disco. Al trabajar con un gran número de datos, estos deben ser cargados en la CPU para ser analizados por la red, es decir, son cargados los tensores en su totalidad a la CPU y el modelo ira extrayendo batches de estos poco a poco. Con los archivos TFRecord, podremos cargarlos todos en el disco y cargar solo los batches con los que vayamos a alimentar a la red en la CPU. Mediante este método y tipo de archivos podremos trabajar con mayores cantidades de datos.

El código que se ha programado se encargara de generar archivos TFRecord de cada modelo, se ejecutara dos veces, una para generar los datos de train y otra para generar los datos de validation.

Para los experimentos venideros se han diseñado 5 datasets distintos, empleando distintas codificaciones.

-Dataset 1: Para cada trayectoria se generarán matrices de dos dimensiones. En la primera dimensión incluyen la codificación del orden en el tiempo y en la segunda dimensión la codificación de la distancia entre saltos sucesivos. Definiremos una N=200.

-Dataset 2: Para cada trayectoria se generarán matrices de tres dimensiones. En la primera dimensión incluyen la codificación del orden en el tiempo, en la segunda dimensión la codificación de la distancia entre saltos sucesivos y en la tercera dimensión donde no se dibujará sobre el esqueleto de la matriz, sino que se rellenaran las 200*200 casillas de la matriz con el valor de la variable máximo tiempo de espera. Definiremos una N=150.

-Dataset 3: Análogamente al dataset 2, se realizará una matriz de 4 dimensiones de profundidad. Las 3 primeras dimensiones serán iguales a las del experimento anterior, añadiremos en la cuarta dimensiones una matriz donde sus 200*200 casillas representen el valor de la desviación estándar del coeficiente de difusión aproximado, la misma codificación que para la tercera dimensión. Empleamos una N=100.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

-Dataset 4: Para cada trayectoria se definirán matrices de dos niveles de profundidad, iguales a los del dataset 1, y dos variables unitarias correspondientes al tiempo máximo de espera y desviación estándar del coeficiente de difusión aproximado, ambas normalizadas con el método comentado anteriormente. Emplearemos una $N=200$, ya que los valores se almacenarán dentro de cada archivo TFrecords como si fuera una variable de profundidad.

-Dataset5: Para cada trayectoria generaremos dos dimensiones iguales a las del dataset 1, solo que modificaremos ligeramente la primera dimensión, correspondiente al orden en el tiempo. Mezclaremos esta codificación con el valor máximo de espera. Para generar esta mezcla sumaremos a cada punto de la trayectoria que no sea de fondo (de 0) el valor de espera máximo de esa trayectoria. Es decir, si por ejemplo trabajamos con una trayectoria con un valor de espera máximo de 27, al normalizarlo sobre 200 como en el experimento 3, tendremos un valor de 0'135 y la trayectoria tiene un total de 100 puntos, que al normalizarlos quedaran del 0 al 1 con pasos de 0'01. Los primeros 5 pasos tendrían un valor de 0'145, 0'155, 0'165, 0'175 y 0'185.

4.2 Diseño del Modelo

Una vez minados y preprocesados los datos, procedemos a trabajar con ellos empleando redes neuronales como las explicadas en el último apartado del marco teórico.

4.2.1 Estrategia de diseño de redes

Programamos una batería de redes con arquitecturas diferentes para entrenarlas simultáneamente y analizar cuál de ellas es mejor para resolver el problema con el dataset generado. Diseñamos la arquitectura de las redes capa a capa.

Estas capas seguirán el orden del modelo teórico explicado en el apartado de teoría. Disponiendo en primer lugar de una fase de extracción de características convolucional y una posterior de clasificación con capas densamente conectadas.

En la fase de extracción de características generaremos cada capa neuronal mediante tres procesos en cadena y en orden. Primero una fase convolucional especificando el tamaño de los filtros, en nuestro caso de 3×3 , y el número de filtros, después una capa de batch normalization especificando el momento, en nuestro caso 0'9, y por último una capa de max pooling, es decir, subsampling basado en máximos, especificando el tamaño de las casillas, en nuestro caso 3.

En la fase de clasificación, cada capa densamente conectada contara con 3 procesos en cadena. Primero una capa de perceptrones, es decir, una capa densa, en la que especificaremos el número de neuronas y la función de activación de estas, en nuestro caso una la función ReLu. Posteriormente una capa de batch normalization con momentum de 0'9 y una última capa de dropout con una ratio de 0'3, es decir, una probabilidad del 30% de apagar las neuronas.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Para diseñar las redes seguiremos un patrón común, definiendo el número de neuronas tanto en las capas convolucionales como densas con potencias de 2, es decir, el número de neuronas de una capa siempre será el doble que el de la anterior. Además de estos, procuraremos que sean potencias de 24 o 16.

Otro factor que tendremos en cuenta en el diseño de la red será la relación de neuronas entre la última capa convolucional y la primera densa, antes y después del flatten, evitaremos compresiones excesivas. Al final de todo el proceso convolucional se generará un número de descriptores total, que corresponderán a el número de neuronas del flatten, por tanto, la primera capa densa deberá de ser suficientemente grande para que la compresión no sea excesiva, esto se daría en el caso de que cada perceptrón de la primera capa densa tuviera demasiados inputs y por tanto un número exagerado de pesos a ajustar.

En el caso de introducir alguna variable de forma directa, es decir, introducir su valor sin pertenecer a ninguna matriz que deba pasar por alguna convolución, esta se introducirá en las capas densas. Emplearemos la función concatenate, la cual añade a cada la capa seleccionada tantas neuronas como variables se añadan, las cuales tendrán ese valor. Si tenemos por ejemplo una capa densa de 16 neuronas y añadimos definimos después de esta la función concatenate con 2 variables externas, la capa pasara a tener 18 neuronas.

De las 150 arquitecturas probadas, de diferentes tamaños y distribuciones, las que mejor resultado nos han dado han sido las siguientes:

Red 1:

- Capas convolucionales: 24, 48, 96, 192 y 384
- Capas densas: 1024, 256, 64 y 16

Red 2:

- Convolucionales: 24, 48, 96, 192 y 384
- Densas: 1024, 256, 64, 16 y 4. Se introducirán las variables unitarias valor de espera máximo y desviación estándar del coeficiente de difusión aproximado justo antes del softmax, en la capa densa de 4 neuronas, teniendo por tanto 6 neuronas en total.

4.2.2 Estrategia de entrenamiento

Entrenamos la red mediante un código train, en el cual llamamos de forma sucesiva todas las redes que hayamos generado en el fichero de networks. Previamente a el entrenamiento de estas definiremos los parámetros optimizador y métrica de la loss, o grafica de perdida.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

El optimizador será el sistema que empleemos para desplazarnos a lo largo de la función gradiente de la función de coste. En nuestro caso hemos empleado el optimizador ADAM (*Adaptative momento estimation*). En él se definen dos momentos que representan la media de los gradientes y la varianza (primer y segundo momento) del gradiente de la función de coste. Además, definimos para este un learning rate de 0'001.

Para la los, o función de perdidas definimos la entropía cruzada o SparseCrossentropy, la cual nos definirá las pérdidas durante el proceso de aprendizaje. Todas las redes serán entrenadas además con 250 epochs.

4.2.3 Métricas de Evaluación

Para evaluar la calidad de nuestros modelos emplearemos tres métricas básicas: precisión, pérdidas (loss) y matriz de confusión.

La **precisión** o accuracy expresa el número de clasificaciones correctas frente a las incorrectas. Debido a que tenemos clases balanceadas, es decir, el mismo número de muestras de todas las clases, podremos decir que esta medida es fiable.

$$\text{Accuracy} = \frac{\text{clasificaciones correctas}}{\text{clasificaciones incorrectas}}$$

Las **perdidas** o loss nos indican como de bien está aprendiendo nuestro modelo, es decir, como se bien se acerca a el mínimo global. Una gráfica de loss estable significara que esta está tendiendo a el mínimo de forma dirigida, una gráfica ruidosa y caótica significara que se está saltando por toda la función gradiente del coste. La loss que empleamos es la entropía cruzada como se ha comentado anteriormente. Para cada clase analiza si la clasificación real coincide con la predicha de cada muestra del batch, es decir, si se acerca al 0 o 1 del softmax, posteriormente calcula el logaritmo del valor de la predicción para cada clase y así deduce si se está acercando o no a el objetivo, ya que esta se reducirá según el valor que se le pase al logaritmo se acerque a 1.

$$\text{Loss}(p, t) = - \sum_{c=1}^c t_{0,c} \log(p_{0,c})$$

Por último, la **matriz de confusión** nos indica la cantidad de muestras que se han clasificado en cada clase, más concretamente, para cada clase, cuantas muestras se han clasificado en la propia clase y cuantas en el resto. La matriz es cuadrada y tiene tantas filas como clases haya, tanto las filas como las columnas representan las clases. La diagonal representa las muestras bien clasificadas y el resto de las posiciones las mal clasificadas. Por ejemplo, si disponemos de una matriz cuadrada de 5×5 , y tenemos en la posición (2,3) el número 23, significara que 23 muestras de la clase 2 han sido clasificadas en la clase 3.

5. Resultados

Mostraremos en este apartado el resultado de los 5 experimentos llevados a cabo, en cada uno de ellos se han probado más de 150 redes, se mostrarán los mejores resultados para cada uno.

5.1 Experimento 1

Empleo de red 1 y dataset 1

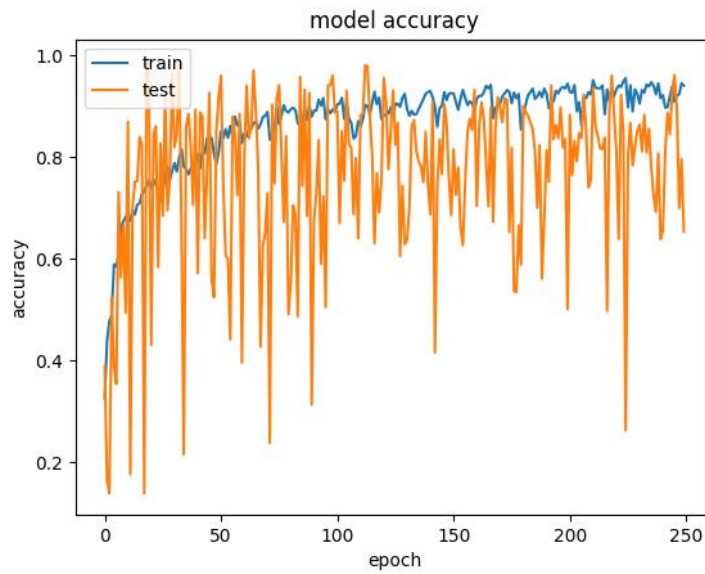


Figura 33: Grafica Accuracy experimento 1

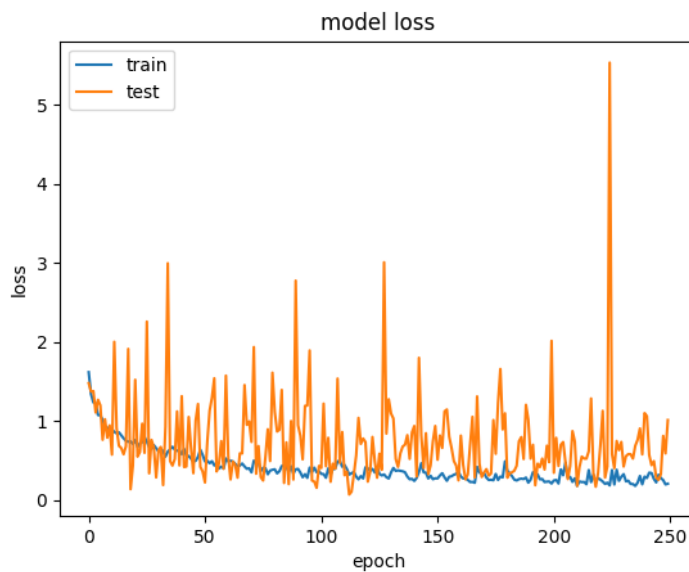


Figura 34: Grafica loss experimento 1

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

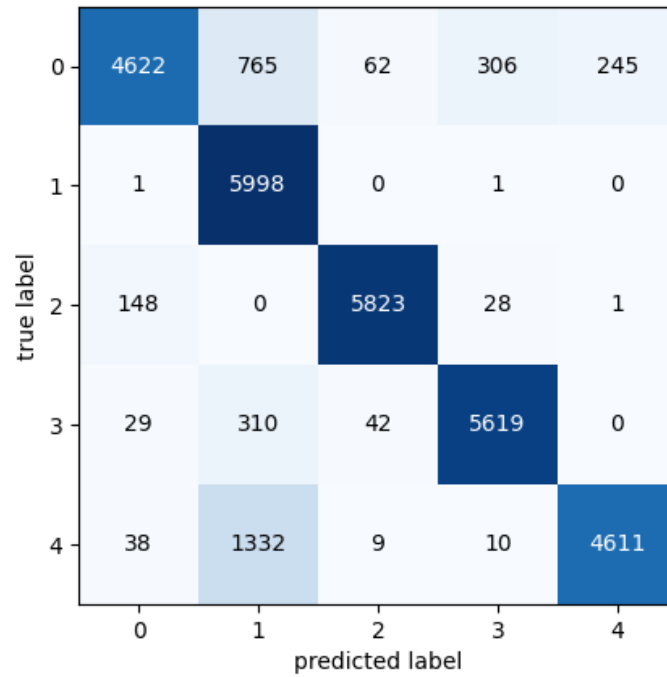


Figura 35: Matriz de Confusión experimento 1

5.2 Experimento 2

Empleo de red 1 y dataset 2

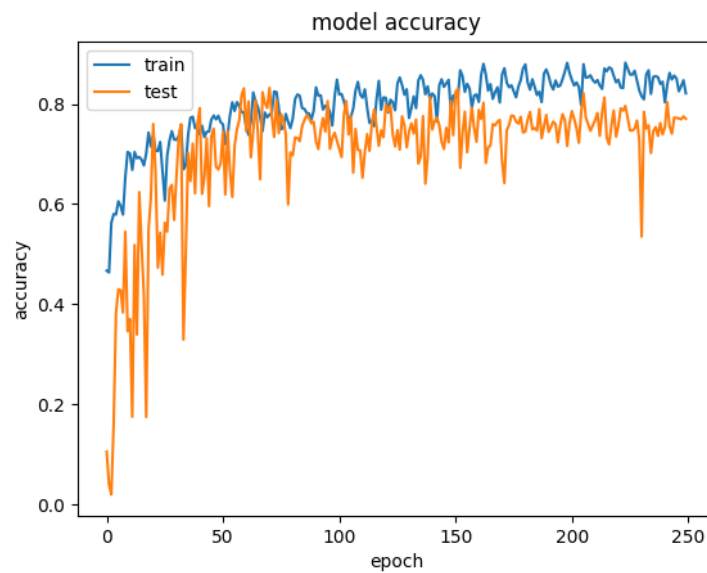


Figura 36: Grafica Accuracy experimento 2

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

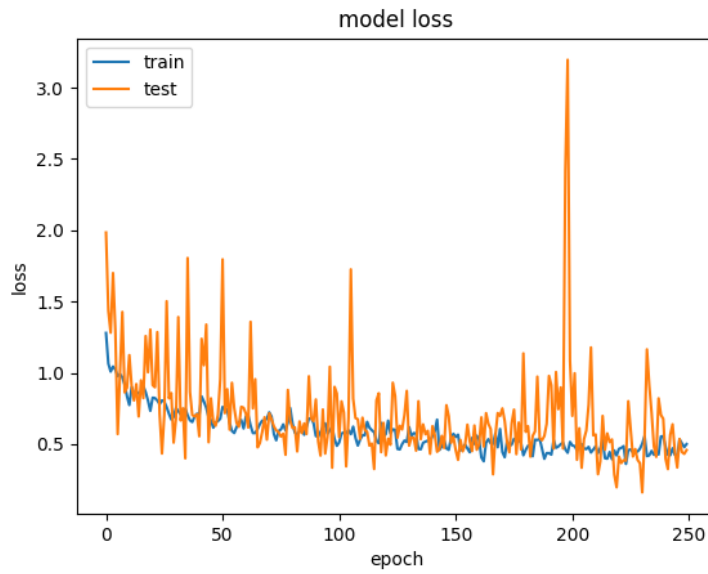


Figura 37: Grafica Loss Experimento 2

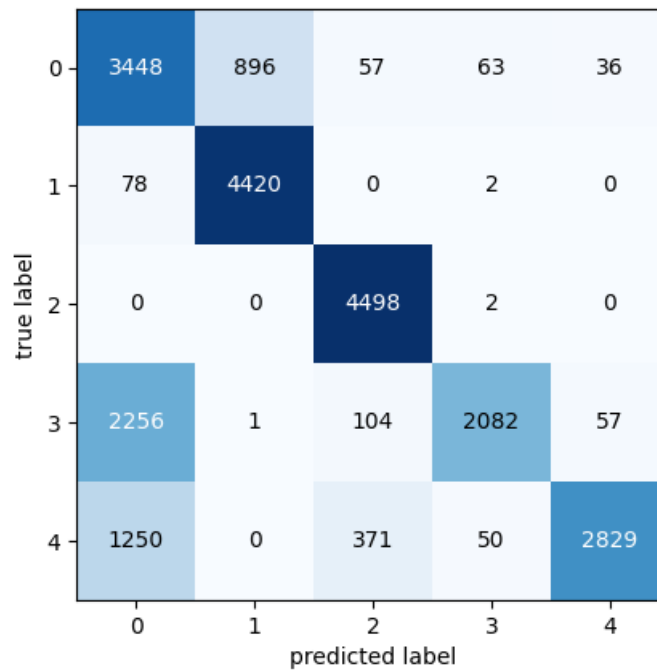


Figura 38: Matriz de Confusión Experimento 2

5.3 Experimento 3

Empleo de red 1 y dataset 3

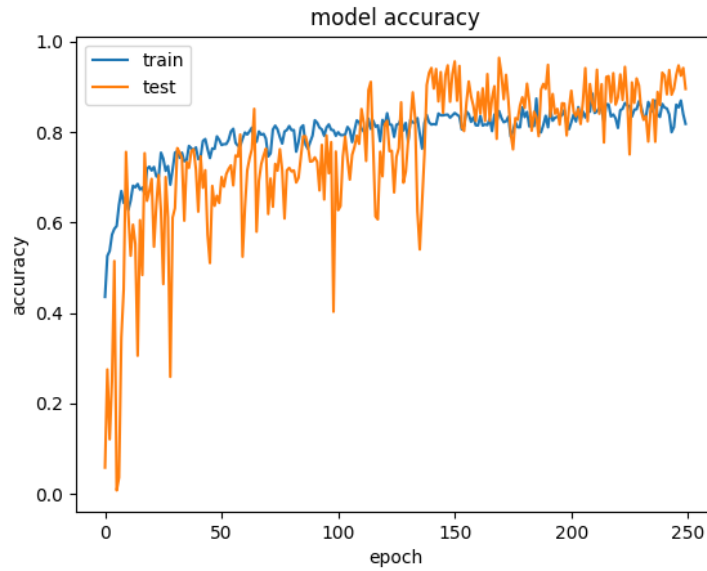


Figura 39: Grafica Accuracy experimento 3

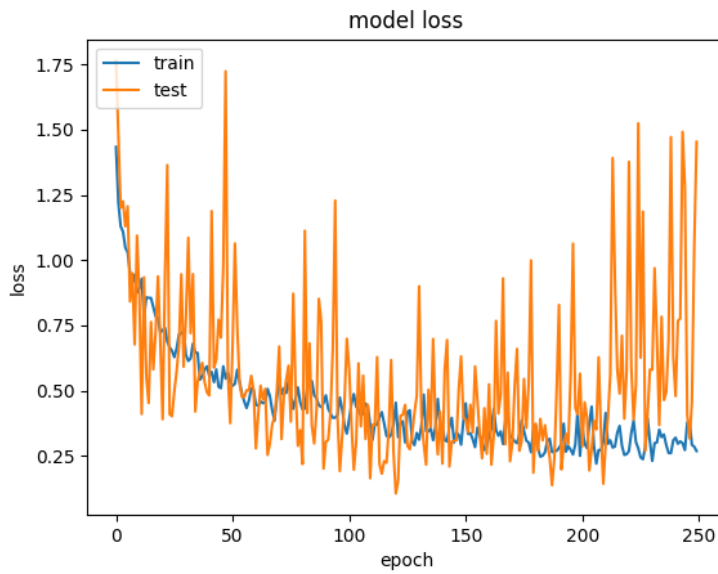


Figura 40: Grafica Loss Experimento 3

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

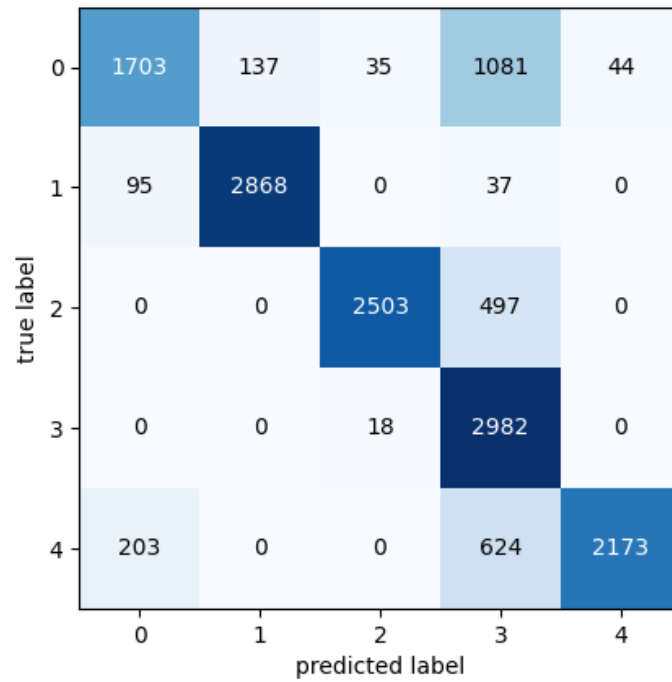


Figura 41: Matriz de Confusión experimento 3

5.4 Experimento 4

Empleo de red 2 y dataset 4

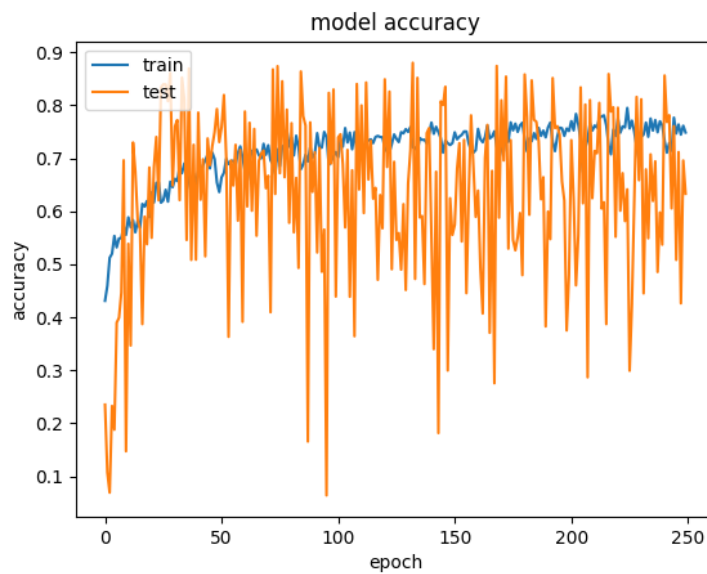


Figura 42: Grafica Accuracy experimento 4

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

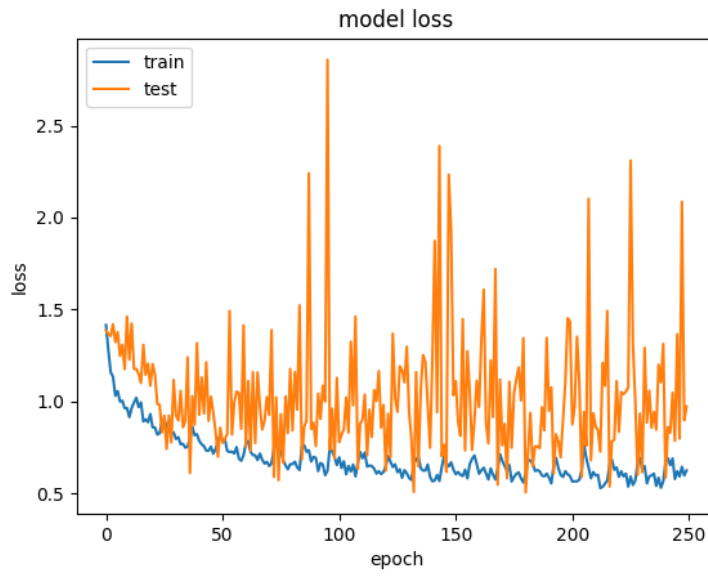


Figura 43: Grafica loss experimento 4

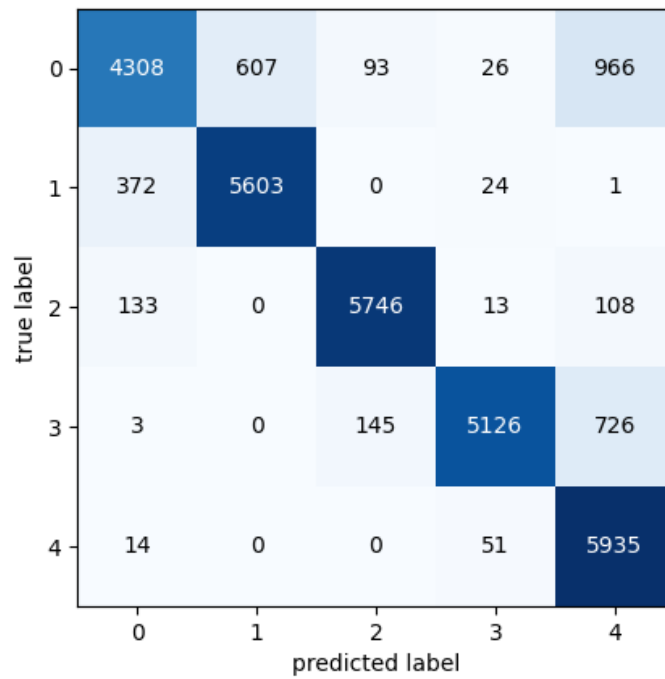


Figura 44: Matriz de Confusión experimento 4

5.5 Experimento 5

Empleo de red 1 y dataset 5

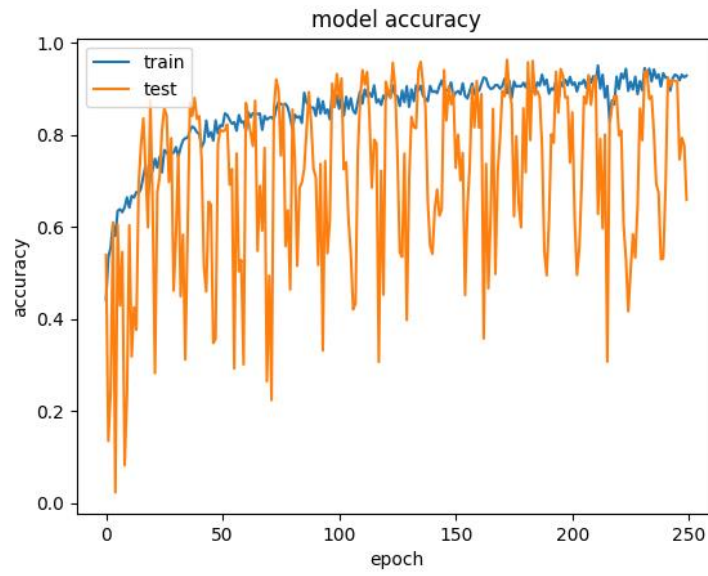


Figura 45: Grafica Accuracy experimento 5

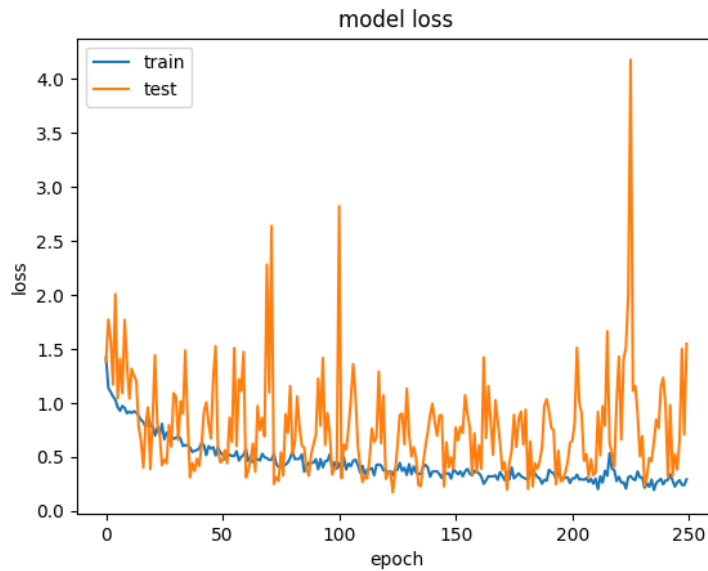


Figura 46: Grafica Loss experimento 5

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

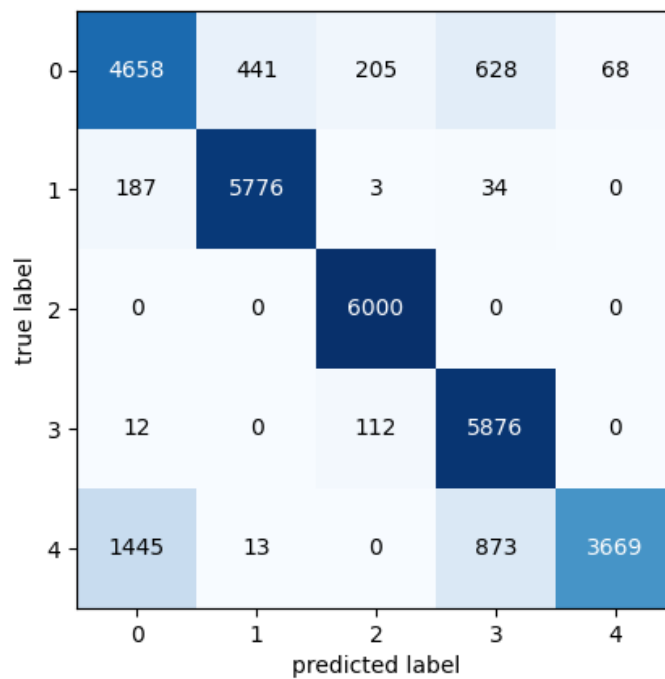


Figura 47: Matriz de confusión experimento 5

6. Discusión

Durante la realización de los experimentos se ha seguido un orden lógico, buscando en cada caso solucionar los problemas del anterior experimento u optimizar su funcionamiento.

En el primer experimento pudimos observar, mediante las métricas empleadas, que los resultados fueron los no deseados, contando con una gran oscilación de la gráfica de accuracy, lo cual nos da unos resultados poco fiables, ya que la precisión no depende del modelo sino del conjunto de validación que empleemos, y en líneas futuras del conjunto que deseemos clasificar. Esto además queda muy patente en la gráfica de pérdidas o loss, la cual presenta grandes oscilaciones, lo cual nos indica que el proceso de aprendizaje no está funcionando correctamente, es decir, la red está moviéndose alrededor de la función de coste de forma prácticamente aleatoria, realizando saltos de punto a punto y algunas veces cayendo en mínimos y otras en superficies muy elevadas y por tanto con mucho error.

Vistos estos resultados consideramos que la red era incapaz de inferir correctos descriptores para diferenciar entre distintos modelos, así pues, decidimos extraer la matriz de confusión para analizar los casos más conflictivos. Como se pudo observar los modelos más conflictivos se tratan del CTRW y el SBM, siendo especialmente problemático el modelo CTRW, el cual se confunde con la mayoría de las modelos presentes, el SBM presenta una elevada confusión con el modelo FBM, debido posiblemente a que ambos se tratan de procesos gaussianos. El resto de las modelos reciben unos scores muy elevados, posiblemente debido a que es muy sencillo inferir descriptores correctos mediante esta codificación, es el caso del LW y FBM, los cuales gracias a sus características espaciales son muy sencillo de detectar, en primer lugar, el LW presenta grandes saltos recorriendo grandes distancias, lo cual queda patente en la segunda dimensión, el FBM presenta grandes tiempos de espera y tiene una morfología muy característica.

Debido a esta problemática, vimos preciso definir nuevos descriptores para poder diferenciar estos dos modelos. Con esto damos inicio al experimento 2.

Definimos por tanto un nuevo descriptor, el tiempo de espera máximo. Este parámetro presenta valores superiores a 300 para el modelo FBM, superiores a 30 para el modelo CTRW y ATTM, y cercanos a 5 para el resto de las modelos. Visto estas características se trata de un parámetro muy interesante, ya que podrá ayudar a la red a reducir la confusión del CTRW con el resto de las modelos, además de otorgar al FBM una nueva característica que ayude a distinguirlo del SBM.

El resultado de este experimento fue muy beneficioso, como era de esperar, aunque presento una nueva problemática, debido a la similitud entre el modelo CTRW y ATTM en cuanto a sus tiempos de espera, este parámetro introdujo ruido indeseado en la discriminación de ambos, como se observa en la matriz de confusión (clases 0 y 3). Debido a este contratiempo, decidimos introducir una nueva variable, una que sea capaz de diferenciar el modelo del resto. Para ello definimos la variable desviación estándar del coeficiente de difusión aproximado, ya que este modelo cuenta con cambios

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

brusco de su coeficiente de difusión, y que no siguen ninguna función de crecimiento como se trata del modelo SBM.

Con esta nueva codificación se realiza el experimento 3, el cual nos da unos resultados muy satisfactorios, consiguiendo reducir la confusión del modelo ATTM con el CTRW, a costa de empeorar ligeramente la confusión del modelo ATTM con el SBM, ya que ambos presentan cambios en su D , y la del modelo CTRW con el ATTM, ya que es capaz de diferenciar perfectamente cuando se trata de ATTM pero en algunos casos (un 35% aproximadamente) los clasifica erróneamente en el ATTM, sigue siendo aun así un mejor resultado. Consiguiendo accuracys del 88% al final del proceso de entrenamiento. Este modelo, aun conservando un muy buen ratio para el LW, lo ha reducido significativamente, ya que en los anteriores experimentos proporcionaba tasas prácticamente perfectas, esto es debido probablemente a la inclusión de nuevos datos, ya que su numero de descriptores óptimos se trataban de los empleados en el anterior experimento.

A modo de optimizar el proceso anterior, se realiza el experimento 4. En este experimento introducimos las variables de una forma mucho más optimizada y elegante, es decir, como neuronas directamente en las capas densas de la fase de clasificación, de esta forma podemos aumentar mucho más el dataset, y así mejorar la precisión de nuestro modelo, e introducir la misma información.

Debido a que nuestra red emplea capas de batch normalization, es decir, normaliza nuestros datos dentro de rangos centrados en media 0 y una desviación típica de ± 1 , debemos de normalizar también nuestros datos, de esta forma se evita meter ruido en la red, ya que introducir datos con valores muy elevados directamente como valores de neuronas, sin pasar por fases de normalizado, juntaría en la misma etapa datos con valores muy dispares entre sí.

Esta normalización la hemos explicado en el apartado de métodos, empleando el método de tukey, debido a que la limpieza de outliers es crucial en este tipo de procesos, ya que una normalización respecto a valores muy descabellados elimina la singularidad del resto de valores que no son exageradamente altos.

De todas las redes que se han empleado, la que mejor resultado ha dado, dentro de lo malo de los resultados, ha sido aquella en la que se introducían los datos en neuronas de las capas finales, y que además cuentan con pocas neuronas, es decir, inmediatamente antes del softmax. Esto intuitivamente tiene mucho sentido y es lo esperable, ya que como se ha visto en el experimento 2, estos parámetros son muy representativos de este tipo de modelos, por tanto, deberán de tener un peso muy elevado como descriptores a la hora de clasificar modelos.

Si introducimos las variables como neuronas en capas muy superficiales, lejos del softmax, y de muchas neuronas, por ejemplo, la capa de 512 neuronas, estos dos parámetros supondrían solo 2 neuronas de un total de 514, bien es cierto que se les daría un peso elevado, pero este se perdería durante las sucesivas etapas, dando mucha más importancia a los descriptores extraídos de las fases convolucionales. Si introducimos estos como hemos realizado, en una última capa de 2 neuronas, tendrá un total actual de 4 neuronas, donde la mitad, es decir, el 50% de la información final será descrita por estos dos parámetros, y los otros 2 de los descriptores extraídos del proceso convolucional.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Este experimento ha arrojado resultados mucho peores que el anterior experimento, el experimento 3, por tanto, deducimos que estas variables no solo ayudan a diferenciar entre parámetros de forma individual, sino que ayudan a reforzar los descriptores durante el proceso de convolución. Como ya se ha comentado anteriormente, la convolución arroja siempre un mapa de una dimensión, independientemente de las dimensiones del input, por tanto, dará el mismo número de descriptores 2 niveles de profundidad que 4, así pues, los dos niveles extras de estas variables reforzaran el proceso de convolución y la extracción de características conjuntamente con los dos primeros niveles.

El experimento 3 por tanto sigue siendo el que mejores resultados nos facilita, pero cuenta con el inconveniente de el gran tamaño de sus datasets, haciendo que tengamos que emplear una $N=100$ en vez de $N=200$. Por tanto, aplicaremos un nuevo experimento con tal de mejorar este factor sin perjudicar al resultado final, este será el experimento 5.

Para realizar este experimento diseñamos un nuevo dataset con el que podamos emplear datasets de $N=200$ y además aprovechar la información de estas variables conjuntamente en el proceso de convolución. Así pues, seleccionamos de las dos variables la más discriminativa, que es el tiempo máximo de espera, porque además de diferenciar la clase CTRW, que es la más problemática como se ha observado en las matrices de confusión, nos permite también diferenciar la clase FBM simultáneamente. Así pues, aplicamos la codificación explicada en el apartado de resultados.

Mediante esta codificación conseguimos mezclar tanto el avance del tiempo como el tiempo de espera máximo, siendo muy discriminativo, ya que aquellas trayectorias con un tiempo de espera más grande tendrán un rango de valores superior en esta dimensión.

El resultado de este experimento contra todo pronóstico ha sido bastante negativo, no se ha conseguido ni de cerca un resultado como los del experimento 3. Esto puede ser debido a los grandes valores del tiempo de espera máximo respecto a los del avance del tiempo (aun normalizados), ya que en el caso del FBM o CTRW, estos dan valores muy elevados respecto a los de la codificación en el tiempo, ocultándola y haciendo que esta pierda todo su significado. Por ejemplo, en el caso de un modelo FBM donde se le asigne un valor 1 al tiempo de espera máximo por haber sobrepasado el umbral de 200, el primer paso corresponderá a un valor de $1'01$, y el siguiente a uno de $1'02$ y sucesivamente, haciendo que los decimales se pierdan y solo se le dé importancia a el valor del tiempo de espera máximo. Por tanto, este tipo de codificación no mezcla satisfactoriamente ambos módulos, consiguiendo la representación de los dos en una misma matriz, sino que enmascara una codificación y solo deja patente una.

Por tanto, el experimento 2 ha sido el más beneficioso, y todas aquellas intenciones de mejorar su codificación con tal de aumentar el tamaño de los datasets han sido un fracaso, ya que no solo debemos fijarnos en la matriz de confusión, sino también en las gráficas de las métricas que empleemos. La presentación como tal de los datos de esa forma ayuda a la red a inferir descriptores útiles para la clasificación de modelos, mediante la combinación de capas durante el proceso de convolución.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN
REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS
CELULARES

Experimento	Diferencian					Tasa de aprendizaje	Accuracy máximo
	CTRW	FBM	LW	ATTM	SBM		
Exp 1	Bien	excelente	excelente	muy bien	Bien	Inestable	90'1%
Exp 2	bien	excelente	excelente	mal	bien	Estable	78'67%
Exp 3	regular	excelente	excelente	excelente	regular	Estable	88'56%
Exp 4	bien	Muy bien	Muy bien	bien	bien	Muy inestable	88'15%
Exp 5	bien	Muy bien	excelente	excelente	bien	Muy inestable	91'07%

Tabla 3: Tabla resumen del comportamiento de cada experimento en la clasificación de cada modelo y su tasa de aprendizaje

Basándonos en (Tabla 3) podemos observar como muchos experimentos, como el experimento 5 o el 1, dan muy buenos resultados, incluso mejores que los del experimento 2 y 3 en cuanto a la clasificación de los modelos. Solo que estos resultados, debido a la inestable tasa de aprendizaje, es decir, tanto su grafica de loss como la de accuracy van dando saltos entre grandes intervalos de valores sucesivamente. Esto nos indica que estos resultados no son fiables, ya que en la época 250 del entrenamiento podría haberse centrado el test en valores de accuracy muy bajos, como por ejemplo del 20%, dándonos así unas clasificaciones erróneas.

7. Conclusión

Basándonos en los objetivos planteados en la introducción consideramos que el trabajo ha cumplido con la mayoría de estos con éxito.

En primer lugar, hemos sido capaces de comprender el problema matemático a un muy alto nivel, desarrollando parte de las fórmulas para entender su explicación (apartado de modelos difusivos), además de aplicar estos conocimientos al desarrollo de las diferentes fases del proyecto.

En segundo lugar, hemos sido capaces de realizar un sistema capaz de clasificar en los diferentes modelos trayectorias reales (diferentes longitudes y niveles de ruido) con una gran precisión según las métricas empleadas, además de mejorar los resultados que disponía hasta ahora la comunidad científica, objetivo por el cual se abrió el *Challenge*.

En tercer lugar, se ha automatizado el proceso mediante algoritmos de Python para facilitar su futuro uso en investigación. Es decir, el programa es capaz de solo con introducirle los datos crudos de las trayectorias (los puntos que ocupan en el espacio), es capaz de realizar sucesivamente el preprocesado y clasificación de estas en los diferentes modelos.

Como inconveniente presentamos la imposibilidad de aportar el score del Challenge en este trabajo, debido principalmente a que el groundtruth se publica el 14 de septiembre y sale de las fechas de presentación de este estudio frente al tribunal.

8. Bibliografía

Cooper, G. M. (2000). *The Cell: A Molecular Approach*. 2nd edition. (2.a ed., Vol. 1). Sunderland (MA): Sinauer Associates. <https://www.ncbi.nlm.nih.gov/books/NBK9898/>

Castelle, J.C & Banfield, J.F. (2018). Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. *Perspective*, Vol. 172, Issue 6, 1181-1197. Doi:10.1016/j.cell.2018.02.016

Van Meer, G., Voelker, D. R., & Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology*, 9(2), 112–124. Doi:10.1038/nrm2330

Megías M, Molist P, Pombal MA. (2020). Atlas de histología vegetal y animal. La célula. Recuperado (fecha de consulta). <http://mmegias.webs.uvigo.es/5-celulas/1-introduccion.php>

Janmey, P. A. & Kinnunen, P. K. J. (2006). Biophysical properties of lipids and dynamic membranes. *Trends in Cell Biology*, 16(10), 538–546. Doi:10.1016/j.tcb.2006.08.009

Branden, C. I. & Tooze, J. (2012). *Introduction to protein structure*. Garland Science. 1-19. https://books.google.es/books?hl=es&lr=&id=eUYWBAAAQBAJ&oi=fnd&pg=PP1&dq=protein+structure&ots=Pab0Lyb1mY&sig=MffAzI_EYI6XsJs33aprqpS_T5M#v=onepage&q=protein%20structure&f=false

von Heijne, G. (2006). Membrane-protein topology. *Nature reviews Molecular cell biology*, 7(12), 909-918. Doi:10.1038/nrm2063

del Campo, J. A. C. F. (1995). Glicoconjugados: Su participación en funciones de los seres vivos. *Mundo científico*, (159), 634-639. <https://dialnet.unirioja.es/servlet/articulo?codigo=225205>

Maldonado Saavedra, O., Ramírez Sánchez, I., García Sánchez, J. R., Ceballos Reyes, G. M., & Méndez Bolaina, E. (2012). Colesterol: Función biológica e implicaciones médicas. *Revista mexicana de ciencias farmacéuticas*, 43(2), 7-22. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1870-01952012000200002

Mukherjee, S., & Maxfield, F. R. (2004). Membrane domains. *Annu. Rev. Cell Dev. Biol.*, 20, 839-866. <https://www.annualreviews.org/doi/abs/10.1146/annurev.cellbio.20.010403.095451>

Simons, K., & Toomre, D. (2000). Lipid rafts and signal transduction. *Nature reviews Molecular cell biology*, 1(1), 31-39. Doi:10.1038/35036052

Dietrich, C., Bagatolli, L. A., Volovyk, Z., Thompson, N. L., Levi, M., Jacobson, K., & Gratton, E. (2001). Lipid rafts reconstituted in model membranes. *Biophysical journal*, 80(3), 1417-1428. Doi:10.1016/s0006-3495(01)76114-0

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Manzo, C., & Garcia-Parajo, M. F. (2015). A review of progress in single particle tracking: from methods to biophysical insights. *Reports on progress in physics*, 78(12), 124601. Doi:10.1088/0034-4885/78/12/124601

Lichtman, J. W., & Conchello, J.-A. (2005). Fluorescence microscopy. *Nature Methods*, 2(12), 910–919. Doi:10.1038/nmeth817

Meddens, M. B. M., de Keijzer, S., & Cambi, A. (2014). High Spatiotemporal Bioimaging Techniques to Study the Plasma Membrane Nanoscale Organization. *Fluorescence Microscopy*, 49–63. Doi:10.1016/b978-0-12-409513-7.00004-x

Magde, D., Elson, E. L., & Webb, W. W. (1974). Fluorescence correlation spectroscopy. II. An experimental realization. *Biopolymers*, 13(1), 29–61. Doi:10.1002/bip.1974.360130103

Betzig, E., Lewis, A., Harootunian, A., Isaacson, M., & Kratschmer, E. (1986). Near Field Scanning Optical Microscopy (NSOM). *Biophysical Journal*, 49(1), 269–279. Doi:10.1016/s0006-3495(86)83640-2

Willig, K. I., Harke, B., Medda, R., & Hell, S. W. (2007). STED microscopy with continuous wave beams. *Nature Methods*, 4(11), 915–918. Doi:10.1038/nmeth1108

Alivisatos, A. P. (1996). Semiconductor Clusters, Nanocrystals, and Quantum Dots. *Science*, 271(5251), 933–937. Doi:10.1126/science.271.5251.933

Fish, K. N. (2009). Total Internal Reflection Fluorescence (TIRF) Microscopy. *Current Protocols in Cytometry*. Doi:10.1002/0471142956.cy1218s50

Klafter J. & Sokolov I.M. (2011). *First Steps in Random Walks: From Tools to Applications*, Oxford University Press. DOI:10.1093/acprof:oso/9780199234868.001.0001

Muñoz-Gil, G., Lewestein, M. & Garcia-March, M.A. (2020). *The random world of moving particles* (Tesis doctoral). Universitat Politècnica de Catalunya, Barcelona.

Guitou, L. & Garcia-March, M.A. (2020). *Single particle Tracking: Detection of anomalous diffusion among living cells by supervised learning methods* (Trabajo Fin de Master). Polytech Nice-Sophia & Universitat Politècnica de València, Valencia.

Wagner, T., Kroll, A., Haramagatti, C. R., Lipinski, H.-G., & Wiemann, M. (2017). Classification and Segmentation of Nanoparticle Diffusion Trajectories in Cellular Micro Environments. *PLOS ONE*, 12(1), e0170165. Doi:10.1371/journal.pone.0170165

Scher, H., & Montroll, E. W. (1975). Anomalous transit-time dispersion in amorphous solids. *Physical Review B*, 12(6), 2455–2477. doi:10.1103/physrevb.12.2455

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Muñoz-Gil, G., Garcia-March, M. A., Manzo, C., Martín-Guerrero, J. D., & Lewenstein, M. (2019). Single trajectory characterization via machine learning. *New Journal of Physics*. doi:10.1088/1367-2630/ab6065

Massignan, P., Manzo, C., Torreno-Pina, J. A., García-Parajo, M. F., Lewenstein, M., & Lapeyre, G. J. (2014). Nonergodic Subdiffusion from Brownian Motion in an Inhomogeneous Medium. *Physical Review Letters*, 112(15). doi:10.1103/physrevlett.112.150603

Lim, S. C., & Muniandy, S. V. (2002). Self-similar Gaussian processes for modeling anomalous diffusion. *Physical Review E*, 66(2). doi:10.1103/physreve.66.021114

Klafter, J., & Zumofen, G. (1994). Lévy statistics in a Hamiltonian system. *Physical Review E*, 49(6), 4873–4877. doi:10.1103/physreve.49.4873

Kowalek, P., Loch-Olszewska, H., & Szwabiński, J. (2019). Classification of diffusion modes in single-particle tracking data: Feature-based versus deep-learning approach. *Physical Review E*, 100(3). doi:10.1103/physreve.100.032410

Thapa, S., Lomholt, M. A., Krog, J., Cherstvy, A., & Metzler, R. (2018). Bayesian nested sampling analysis of single particle tracking data: maximum likelihood model selection applied to stochastic diffusivity data. *Physical Chemistry Chemical Physics*. doi:10.1039/c8cp04043e

Bo, S., Schmidt, F., Eichhorn, R., & Volpe, G. (2019). Measurement of anomalous diffusion using recurrent neural networks. *Physical Review E*, 100(1). doi:10.1103/physreve.100.010102

Zhou, Q., Shen, Z., Li, S., Li, X., Wang, W., & Wang, Z. (2008). Robust and traditional statistical methods in the establishment of immunoglobulin E target values in external quality assessment program. *Clinica Chimica Acta*, 387(1-2), 66–70. doi:10.1016/j.cca.2007.09.003

Bagnato, J.I. (2018). Aprende Machine Learning [Mensaje en un blog]: ¿Cómo funcionan las Convolutional Neural Networks? Visión por Ordenador. Recuperado de: <https://www.aprendemachinelearning.com/como-funcionan-las-convolutional-neural-networks-vision-por-ordenador/>

Garcia Ferrando, G., Garcia-Gomez, J.M. & Monserrat, C. (2017). Deep learning en segmentación de imagen medica (Trabajo final de grado). Universitat Politècnica de València, Valencia.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Lizana Nuñez, M. (2015). Liferder [Mensaje en un blog] : Membrana Plasmática: Características, Funciones y Estructura. Recuperado de: <https://www.liferder.com/caracteristicas-membrana-plasmatica/>

Lison, P. (2018). Bioquímica [Asignatura]. Universitat Politècnica de València (UPV), Valencia.

Membrana Plasmática (2020). En Wikipedia: https://es.wikipedia.org/wiki/Membrana_plasm%C3%A1tica_eucariota

CajalseyGalileos (2013). Como secuestran los virus la maquinaria endocítica? (2) Nature. Recuperado de: <https://cajalseygalileos.wordpress.com/2013/04/26/como-secuestran-los-virus-la-maquinaria-endocitica-ii-nature/>

Universidad de alcala (UAH), (2011). Biomodel [Mensaje en un blog]: Balsas de membrana (membrane rafts). Recuperado de: <http://biomodel.uah.es/model2/lip/lipid-rafts.htm>

Fogarty, K. H., Zhang, W., Grigsby, I. F., Johnson, J. L., Chen, Y., Mueller, J. D., & Mansky, L. M. (2011). New Insights into HTLV-1 Particle Structure, Assembly, and Gag-Gag Interactions in Living Cells. *Viruses*, 3(6), 770–793. doi:10.3390/v3060770

Glen, S. (2016). Heavy Tailed Distribution & Light Tailed Distribution: Definition & Examples. Recuperado de: <https://www.statisticshowto.com/heavy-tailed-distribution/>

Gimp.org (2016). Convolutional Matriz [Post]. Recuperado de: <https://docs.gimp.org/2.8/en/plugin-convmatrix.html>

Sharma, S. (2017). Towards data Science [Mensaje en un blog]: Activation Functions in Neural Networks. Recuperado de: <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6>

Rahman, N. (2019). Quora [Mensaje en un blog]: What is the benefit of using average pooling rather than max pooling?. Recuperado de: <https://www.quora.com/What-is-the-benefit-of-using-average-pooling-rather-than-max-pooling>

Torres, J. (2018). Jordi TORRES.AL [Mensaje en un blog]: Deep learning. Introducción practica con keras (primera parte). Recuperado de: <https://torres.ai/deep-learning-inteligencia-artificial-keras/>

Educative.io, (2017). Edpresso [mensaje en un blog]: learning rate in machine learning. Recuperado de: <https://www.educative.io/edpresso/learning-rate-in-machine-learning>

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS BROWNIANOS ANÓMALOS BASADO EN REDES CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y SEGMENTACIÓN EN ENTORNOS CELULARES

Dubbura, I. (2018). Towards Data science [Mensaje en un blog]: Coding neural network-Dropout. Recuperado de: <https://towardsdatascience.com/coding-neural-network-dropout-3095632d25ce>

Shreyak, J. (2020). Becoming Human [Mensaje en un blog]: Building a Convolutional Neural Network (CNN) Model for Image classification. Recuperado de: <https://becominghuman.ai/building-a-convolutional-neural-network-cnn-model-for-image-classification-116f77a7a236>

PRESUPUESTO

DISEÑO DE SISTEMA DE CLASIFICACIÓN DE MOVIMIENTOS
BROWNIANOS ANÓMALOS BASADO EN REDES
CONVOLUCIONALES PARA SU CARACTERIZACIÓN Y
SEGMENTACIÓN EN ENTORNOS CELULARES

AUTOR: ALBERT BELENGUER I LLORENS

TUTOR: JUAN MIGUEL GARCÍA GÓMEZ

COTUTORES: MIGUEL ÁNGEL GARCÍA-MARCH Y JAVIER JUAN ALBARRACÍN

CURSO: 2019-2020

ÍNDICE DE PRESUPUESTO

Introducción.....	1
Coste Hardware.....	2
Coste Software.....	2
Coste Personal.....	3
Total.....	3

Introducción

Dividiremos el presupuesto en 3 secciones. En la primera, el hardware, en la cual contaremos todos los materiales que se han empleado para el proyecto. En la segunda, el software, es decir, el coste de los soportes digitales que hemos empleado para dar soporte al hardware. Por último, el coste del personal, en esta sección se dispondrán los salarios de todos los participantes directos del proyecto, es decir, aquellos que han colaborado activamente en el desarrollo del estudio. Para el desarrollo total del proyecto se han estimado un total de 400 horas de trabajo, distribuidas en 6 meses de trabajo.

Para el desarrollo de la sección de hardware se ha tenido en cuenta la amortización de los materiales, estimando de esta forma los años de vida útil de las máquinas empleadas y así pudiendo calcular el precio estimado que ha supuesto en horas su uso.

Para el desarrollo del apartado de software, análogamente a el apartado de hardware, se ha calculado la amortización de cada licencia. En cierto modo, las licencias empleadas han sido suministradas por la UPV (costando al proyecto un coste de 0 €). Aunque contaremos el coste de este debido a que es un coste que tiene que costear la UPV

Para el apartado de Coste de personal se ha considerado un suelo basado en los estándares de los diferentes tipos de personal. Siendo de 20.000 € al año para el ingeniero biomédico junior, de unos 40.000€ para el doctor en ingeniería y 50.000€ para el catedrático en ingeniería. Estos sueldos se tratan del salario bruto, es decir, hay que sumar el coste de la seguridad social por parte de la empresa (en este caso la UPV). Este coste se calcula como un porcentaje del salario bruto total, desglosándolo según indica el ministerio de trabajo, migraciones y seguridad social: 23'6% como cotización de contingencias comunes, 5'5% como cotización por paro, 1% como cotización de accidentes y enfermedades, 0'6% por formación profesional y 0'2% como cotización de fondos de garantía salarial (FOGASA), todo esto resulta en un total del 30'9% sobre el salario bruto. El total de horas realizadas al año han sido consideradas como 1780 h, considerando un total de horas máximas de trabajo diario de 8 h, y restándole los días que caen en fin de semana, 15 días festivos y 30 días de vacaciones.

Coste hardware

Unidad	Descripcion	Detalles	Cantidad (u)	Precio (€)	Amortizacion (años)	Total (€)
u	PC Lenovo Z51	Procesador: Intel® Core™ i5-5200U CPU @ 2.20GHz (4 CPUs), - 2.2GHz Tarjeta grafica: AMD Radeon™ R9 M375 RAM: 12288 MB Microsoft Windows 8 incluido	1	756	5	75'5
u	Cluster unidad de bioinformatica y bioestadística Instituto Principe Felipe	44 nodos con mas de 660 CPU y memoria RAM de 11 TeraB. Capacidad de 1 PetaB	1	50000	10	2500
Total:						2575'5

Coste Software

Unidad	Descripcion	Detalles	Cantidad (u)	Precio (€)	Amortizacion (años)	Total (€)
licencia	Microsoft Office 2016	Paquete office que incluye licencias de Word, powerpoint, teams, outlook, oneNote, etc	1	399'99	1	399'99
licencia	Python	Lenguaje de programacion Python y entorno de programación Anaconda	1	0	0	0
Total:						399'99

Coste Personal

Descripcion	Tarea	Nombre	Rango	Total (horas)	Precio por hora (€)	Seguridad social (€)	Salario (€)	Coste totales (€)
Ingeniero biomédico	Desarrollo del proyecto	Albert Belenguer i Llorens	Junior	400	14'7	1816'92	4063'08	5880
catedrático en informática	Supervision y direccion del proyecto	Juan Miguel Gracia Gómez	Senior	100	36'76	1135'88	2540'11	3676
Doctor en Ingeniería Industrial	Soporte en tareas relacionadas con la matemática y física	Miguel Angel García-March	Senior	100	29'41	908'76	2032'24	2941
Doctor en informática	Soporte y desarrollador de algoritmos y tareas informáticas	Javier Juan Albarracín	Senior	100	29'41	908'76	2032'24	2941
Total:								15438

Total

Descripcion	Coste (€)
Coste Hardware	2575'5
Coste Software	399'99
Coste Personal	15438
Total:	18413'49