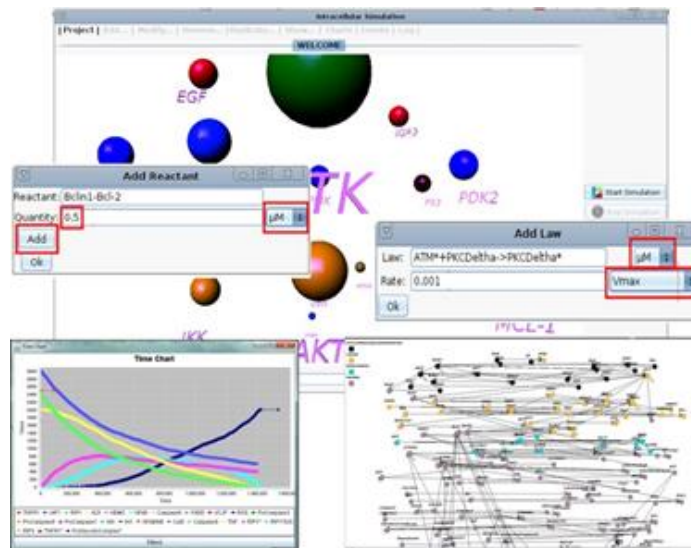


Proyecto académico y programa de actividades a desarrollar durante el período sabático comprendido del 1 de octubre de 2025 al 30 de noviembre de 2026 (14 meses)

Dr. Pedro Pablo González Pérez

**Departamento de Matemáticas Aplicadas y Sistemas
División de Ciencias Naturales e Ingeniería
UAM, Unidad Cuajimalpa**

“Biología Computacional y Bioinformática: Modelado, Simulación, y Análisis de Datos a Gran Escala”



Contenido

1. Introducción	3
1.1. Biología computacional y bioinformática.....	3
1.2. Datos a gran escala.....	4
1.3. Tipos de datos a gran escala atendiendo su naturaleza	5
1.3.1. Datos reales.....	5
1.3.2. Datos producidos por la simulación computacional	5
1.3.3. Datos sintéticos: datos generados de forma artificial	6
1.4. La simulación computacional: la plataforma bioinformática Cellulat.....	6
2. Objetivos del proyecto	8
2.1. Objetivo general.....	8
2.2. Objetivos específicos.....	8
3. Productos esperados.....	9
4. Participantes, instituciones receptoras y estancias de trabajo.....	9
5. Enfoque metodológico y etapas de desarrollo	10
6. Vinculación con los proyectos de investigación, planes y programas académicos de la UAM, Unidad Cuajimalpa	14
7. Referencias.....	15

1. Introducción

Durante las últimas dos décadas, la biología celular y molecular se ha visto ampliamente favorecida por el desarrollo tecnológico en áreas de la computación tales como hardware, software, comunicaciones, y tecnologías de bases de datos, lo cual ha conllevado al desarrollo y auge de dos grandes escenarios:

- 1) Por una parte, la simulación computacional en ciencias biológicas y biociencias, que más allá de proporcionar un entorno virtual para la experimentación *in silico*, ha sido capaz de generar enormes volúmenes de datos de calidad, listos para su ulterior análisis y uso en la toma de decisiones.
- 2) Por otra parte, el desarrollo de las ciencias ómicas, tales como genómica, proteómica, transcriptómica, y metabolómica, lo cual ha conllevado a la producción masiva de datos biológicos, también listos para ser analizados e interpretados.

Ambos escenarios han dado paso a la era de la producción de los datos a gran escala –del inglés, *big data*– (Aguilar, 2013; Marr, 2015; Mayer-Schönberger & Cukier, 2013, 2017) en la biología y ciencias biológicas, evidenciándose un aumento exponencial en la generación de datos lo cual, como consecuencia, ha conllevado al desarrollo de métodos y herramientas que permitan el almacenamiento, procesamiento y análisis eficiente de estos datos masivos (Jabalia & Lakshmi, 2020; Ohyanagi et al., 2022; Whitlock & Schluter, 2014; Ye, 2021).

1.1. Biología computacional y bioinformática

Los múltiples y novedosos resultados provenientes en los últimos años de campos de las ciencias de la computación, ciencias de la información, matemáticas aplicadas y la ingeniería se han encontrado con la biología, guiándonos a una nueva área, emergiendo disciplinas tales como:

- Bioinformática
- Biología computacional

Por una parte, la bioinformática puede definirse como el desarrollo sistemático y la aplicación de métodos, técnicas y modelos computacionales (sistemas y soluciones computacionales) al análisis de datos biológicos, provenientes de fuentes como la investigación básica, el modelado, las búsquedas en bases de datos biológicas, etc. (Baxevanis et al., 2020; Compeau & Pevzner, 2018; Ewens & Grant, 2004; Lesk, 2019; Zvelebil & Baum, 2007) De esta forma, la bioinformática se enfoca en el desarrollo y aplicación de técnicas y métodos informáticos para procesar y analizar datos biológicos, en particular aquellos provenientes de tecnologías relacionadas con la proteómica, genómica, transcriptómica, metabólica, etc.

Por otra parte, la biología computacional, estrechamente relacionada con la bioinformática, se centra en el desarrollo y aplicación de modelos computacionales y simulaciones para estudiar sistemas biológicos. En otras palabras, la biología computacional abarca el estudio, comprensión, modelado y simulación a nivel sistema de los sistemas biológicos, considerando las interacciones complejas que ocurren entre sus componentes elementales (por ejemplo, proteínas, genes y elementos celulares) (Alon, 2019; Kelley & Didulo, 2018; O'Neil, 2017; Waterman, 1995). Siguiendo este enfoque, un modelo comprensible del sistema completo podría ser desarrollado a partir del estudio de las interacciones y relaciones entre los componentes y partes del sistema biológico (por ejemplo, redes metabólicas, redes genéticas, células, organismos, etc.).

1.2. Datos a gran escala

El término "datos a gran escala" o "datos masivos" ha ganado gran popularidad en el contexto actual de la tecnología de la información para referirse a los enormes y complejos conjuntos de datos producidas por múltiples fuentes digitales. Cuando pensamos en datos a gran escala seguramente estamos describiendo problemas que surgen con relación a (Aguilar, 2013; Marr, 2016; Marz & Warren, 2015; Mayer-Schönberger & Cukier, 2013, 2017):

- Datos cuyo volumen y complejidad requieren métodos más sofisticados de almacenamiento, recuperación, interacción, preparación, análisis-inferencia y presentación.
- Grandes volúmenes de datos que implican tanto datos estructurados como no estructurados, lo que hace que su tratamiento sea mucho más complejo.
- La aplicación de un potente procesamiento computacional a conjuntos de datos altamente masivos y complejos.

En los últimos años, la definición de datos a gran escala ha abarcado de tres a cinco dimensiones clave, a las que comúnmente se hace referencia como las "tres V" o las "cinco V". Es decir, la definición original caracterizada por volumen, velocidad y variedad; o la definición ampliada, al incluir veracidad y valor (Marz & Warren, 2015; Mayer-Schönberger & Cukier, 2017; Tolk, 2015).

Las siguientes son algunas de las principales fuentes generadoras de datos a gran escala, entre las que destacan la simulación computacional, y la bioinformática y la biología computacional:

- Comercio electrónico (e-commerce)
- Internet, motores de búsqueda y redes sociales
- Mercado financiero
- Datos producidos por sensores

- Simulaciones computacionales
- Bioinformática y Biología Computacional
- Técnicas de generación de datos sintéticos
- Sistemas computarizados para el cuidado y monitoreo de la salud
- Dispositivos móviles y aplicaciones (apps)
- Plataformas digitales de entretenimiento

1.3. Tipos de datos a gran escala atendiendo su naturaleza

Atendiendo a su naturaleza, los datos a gran escala pueden ser clasificados en:

- Datos reales
- Datos generados por la simulación computacional
- Datos sintéticos

1.3.1. Datos reales

Las observaciones o mediciones tomadas directamente del mundo real, se denominan datos reales o genuinos. Los datos reales comúnmente provienen de estudios de investigación, encuestas, trabajo de campo, lecturas de sensores que documentan sistemas o procesos reales, plataformas de comercio electrónico, plataformas del mercado financiero, etc. Los datos reales son valiosos porque reflejan la realidad y brindan información auténtica sobre los eventos y procesos que se están estudiando. Sin embargo, en algunos casos, obtener datos reales puede ser costoso, llevar mucho tiempo o incluso ser imposible debido a limitaciones éticas o logísticas.

1.3.2. Datos producidos por la simulación computacional

En las últimas tres décadas, la simulación computacional y la experimentación in silico se han extendido a través de la gran mayoría de las disciplinas científico-tecnológicas, destacándose entre éstas ciencias de la vida, biología celular y molecular, química, física, ciencias ambientales, y varias áreas de la ingeniería, entre otras. Los grandes volúmenes de datos producidos por la simulación computacional constituyen un valioso insumo que fomenta el avance y la innovación en las áreas antes mencionadas.

Los datos producidos por simulaciones computacionales se generan a partir de la ejecución de modelos matemáticos, algoritmos, o técnicas de inteligencia artificial en un entorno virtual (comúnmente, software y hardware) (Lechler et al., 2021; Sánchez-Gutiérrez & González-Pérez, 2022). Los datos generados representan resultados teóricos o hipotéticos basados en los supuestos y parámetros de la simulación (Tolk, 2015). La simulación computacional permite el análisis predictivo y ofrece información útil sobre cómo un sistema puede reaccionar en diversas

circunstancias. La confiabilidad de los modelos e hipótesis subyacentes determina la precisión y aplicabilidad de los datos de simulación.

1.3.3. Datos sintéticos: datos generados de forma artificial

El término "datos sintéticos" se refiere a conjuntos de datos que se han producido artificialmente pero que se parecen mucho a los datos reales. La capacidad de los datos sintéticos para complementar conjuntos de datos pequeños o desequilibrados y mejorar los modelos de aprendizaje automático de clasificación los ha hecho populares en muchos campos diferentes (Abufadda & Mansour, 2021; Goyal & Mahmoud, 2024; Hernandez et al., 2022; Murtaza et al., 2023). Cuando los datos reales son difíciles de conseguir, confidenciales o raros, los datos sintéticos pueden ser especialmente útiles.

Los datos sintéticos se generan utilizando varias técnicas de generación de datos, como las redes generativas adversarias (GAN), la técnica estadística de sobremuestreo de minorías sintéticas (SMOTE), o algoritmos de interpolación. Los modelos de aprendizaje automático, la validación de algoritmos y la ampliación de conjuntos de datos se pueden realizar utilizando datos sintéticos.

1.4. La simulación computacional: la plataforma bioinformática Cellulat

El objetivo de la plataforma de simulación computacional Cellulat es proporcionar un laboratorio virtual bioinformático para el desarrollo de la experimentación in silico en sistemas de señalización celular, caracterizada por la robustez, precisión y flexibilidad. El ambiente de experimentación in silico soportado por la plataforma Cellulat incluye herramientas para la simulación, exploración, análisis y predicción de este tipo de sistema biológico.

La herramienta computacional Cellulat (ver figura 1.1.) constituye un robusto y flexible laboratorio virtual bioinformático dedicado a la simulación, exploración, análisis y predicción de sistemas moleculares a nivel celular, en particular, sistemas de señalización celular (Cárdenas-García et al., 2016; Cardenas-Garcia & González-Pérez, 2018; González Pérez et al., 2013; Ortiz-González et al., 2022). Cellulat proporciona al usuario (biólogo, bioquímico, investigador en el área biomédica, ciencias de la vida, etc.) el soporte necesario para el diseño de experimentos in silico que permitan:

- 1) Modelar, simular y visualizar de forma integral la complicada estructura y el funcionamiento de las vías de señalización celular.
- 2) Visualizar (a través de gráficos, curvas concentración/tiempo y tablas) el estado de actividad de la red general, las interacciones que ocurren entre los diferentes elementos de señalización y las variaciones de sus estados (concentración, actividad, etc.) en el tiempo.

- 3) Predecir los efectos a nivel celular (por ejemplo, proliferación, muerte celular, apoptosis, etc.) de los cambios y perturbaciones en el sistema en tiempo real y en las condiciones fisiológicas requeridas. Tales cambios y perturbaciones incluyen variación en la concentración molar de los elementos de señalización, mutaciones a proteínas, incorporación de otros elementos de señalización en la red de señalización, eliminación total de elementos de señalización (tipo knock-out virtual).

La nueva funcionalidad a incorporar en la plataforma bioinformática Cellulat será:

- Producción, pre-procesamiento y registro de grandes volúmenes de datos (big data/data farming) resultantes de la ejecución de la simulación y experimentos in silico, para su posterior uso en tareas de data mining, deep learning, etc.

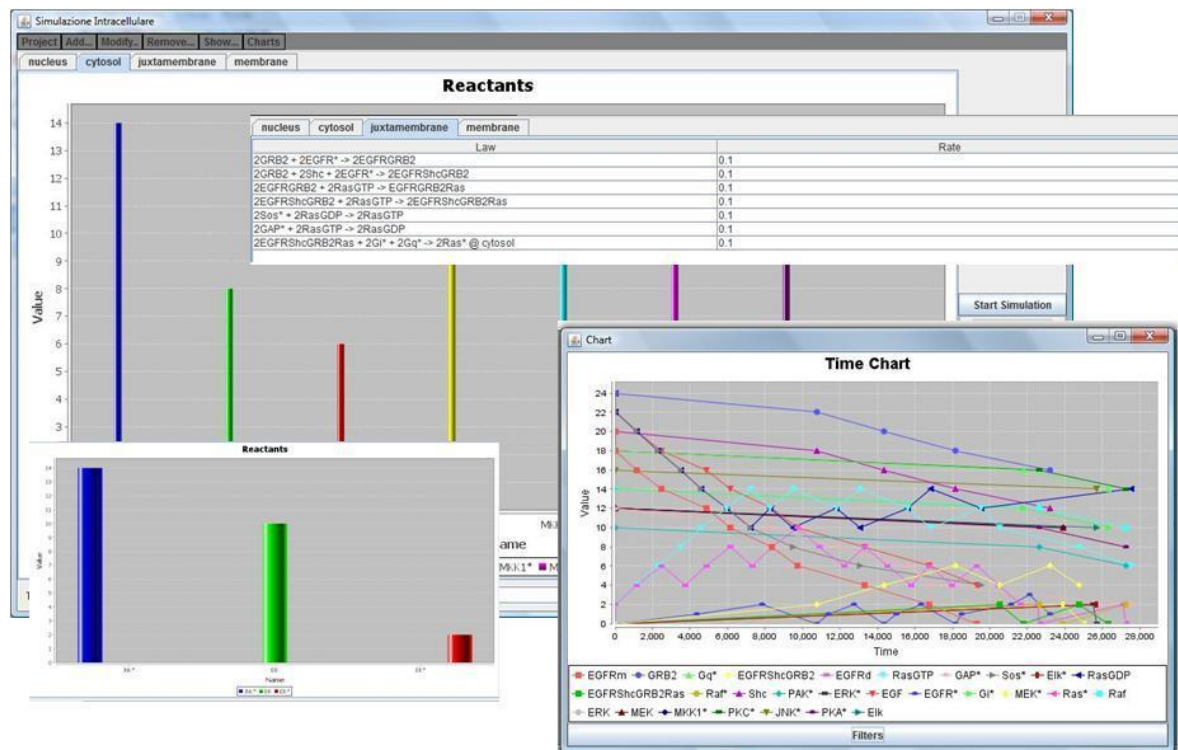


Figura 1.1. La plataforma bioinformática Cellulat para la simulación y experimentación in silico de redes de señalización celular.

2. Objetivos del proyecto

2.1. Objetivo general

Fortalecimiento de la red académica “Computational Biology: Modeling, Simulation, Big Data, Data Mining and Machine Learning”, a través de actividades conjuntas en investigación, desarrollo de software, docencia, y divulgación. La red académica está integrada por las siguientes instituciones:

- Universidad Autónoma Metropolitana, Unidad Cuajimalpa, Departamento de Matemáticas Aplicadas y Sistemas, Ciudad de México, México
- Universidad Autónoma de la Ciudad de México, Colegio de Ciencia y Tecnología – Cuauhtémoc, Ciudad de México, México
- Benemérita Universidad Autónoma de Puebla, Facultad de Medicina, y Herbario y Jardín Botánico Universitario, Puebla, México
- Università degli Studi di Bologna, Campus di Cesena, Dipartimento di Informatica - Scienza e Ingegneria, Cesena, Italia

2.2. Objetivos específicos

1. Incrementar y perfeccionar la simulación computacional en bioinformática con la generación de datos a gran escala. Añadir la nueva funcionalidad “generación de datos a gran escala” a la plataforma bioinformática *Cellulat* (<http://bioinformatics.cua.uam.mx/site/>), a través de los procesos de ingeniería de software inversa y progresiva.
2. Desarrollo de experimentación in silico en la nueva plataforma bioinformática *Big Data - Cellulat*, con la finalidad de producir grandes volúmenes de datos de calidad para su posterior análisis con técnicas de machine learning.
3. Desarrollar herramientas de minería de datos que puedan abordar los dos grandes problemas que comúnmente se presentan en los grandes volúmenes de datos biológicos y biomédicos: características redundantes y desbalance de clases. Concluir el desarrollo de la herramienta de minería de datos *IDA WEB TOOL* (<https://garciad955.github.io/IW-IDAT/>).
4. Efectuar la minería de datos de los volúmenes de datos producidos por la plataforma bioinformática *Big Data - Cellulat*, utilizando la funcionalidad soportada por la herramienta de minería de datos *IDA WEB TOOL*.
5. Propuesta y ensayo de un enfoque metodológico para abordar el problema del desbalance de clases en conjuntos de datos a gran escala, a partir de la generación de datos sintéticos basada en técnicas de inteligencia artificial generativa

3. Productos esperados

En la tabla 3.1. se relacionan los productos esperados resultantes del período sabático, clasificados en las categorías investigación, docencia, desarrollo, y divulgación.

Tabla 3.1: Productos esperados del periodo sabático

Producto esperado	Clasificación del producto
Versión actualizada de la plataforma bioinformática <i>Cellulat</i> , con la capacidad añadida de generación de grandes volúmenes de datos (<i>data farming</i>), como resultado de la ejecución de la simulación	Desarrollo/Investigación
Versión actualizada de la herramienta computacional de minería de datos <i>IDA WEB TOOL</i>	Desarrollo/Investigación
Manuscrito del libro de texto intitulado “Biología Computacional y Bioinformática: Un enfoque práctico”, como soporte a las UEA Bioinformática I y Bioinformática II del Posgrado en Ciencias Naturales e Ingeniería	Docencia
Manuscrito del artículo de investigación intitulado “Dealing with Medical Imbalanced Datasets in Supervised Machine Learning: An Approach Based on Augmentation and Interpolation Techniques”	Investigación
Manuscrito del artículo de investigación intitulado “Increasing computer simulation in molecular biology with data farming and synthetic data: a case study in cell signaling networks”	Investigación
Manuscrito del artículo relacionado con el modelado y simulación de las vías de señalización asociadas a atrofia e hipertrofia muscular	Investigación
Participación en un evento académico internacional (congreso, conferencia, o seminario)	Investigación/Divulgación
Participación en un evento académico nacional (congreso, conferencia, o seminario)	Investigación/Divulgación
Presentación de seminarios de investigación a nivel nacional e internacional	Docencia/Investigación/Divulgación

4. Participantes, instituciones receptoras y estancias de trabajo

Los participantes e instituciones receptoras del periodo sabático son aquellas que conforman la red académica “Computational Biology: Big Data, Data Mining and Machine Learning”, que se indican a continuación:

- Universidad Autónoma Metropolitana, Unidad Cuajimalpa, Departamento de Matemáticas Aplicadas y Sistemas, Ciudad de México, México
 - ❖ Dr. Pedro Pablo González Pérez,
email: pgonzalez@cua.uam.mx
- Universidad Autónoma de la Ciudad de México, Colegio de Ciencia y Tecnología – Cuauhtémoc, Ciudad de México, México
 - ❖ Dr. Máximo Eduardo Sánchez Gutiérrez
email: maximo.sanchez@uacm.edu.mx
- Benemérita Universidad Autónoma de Puebla, Facultad de Medicina y Herbario y Jardín Botánico Universitario
 - ❖ Dra. Maura Cárdena García
maura.cardenas@correo.buap.mx
 - ❖ Dra. María Guadalupe Hernández Linares
guadalupe.mghl@correo.buap.mx
- Università degli Studi di Bologna, Campus di Cesena, Dipartimento di Informatica - Scienza e Ingegneria, Italia.
 - ❖ Dr. Andrea Omicini
email: andrea.omicini@unibo.it
 - ❖ Dr. Alessandro Ricci
a.ricci@unibo.it
 - ❖ Dra. Sara Montagna
sara.montagna@unibo.it

El período sabático abarcará diferentes estancias iterativas de trabajo en las tres instituciones receptoras: Università degli Studi di Bologna, Campus di Cesena, Dipartimento di Informatica - Scienza e Ingegneria, Italia; Universidad Autónoma de la Ciudad de México, Colegio de Ciencia y Tecnología – Cuauhtémoc, Ciudad de México, México; Benemérita Universidad Autónoma de Puebla, Puebla, México.

El tiempo en semanas o meses que se dedicará a cada una de estas estancias se irá definiendo sobre la marcha, según los avances y exigencias del propio proyecto sabático. Sin embargo, se estima que las estancias que abarcarán más tiempo serán las que se llevarán a cabo en la institución receptora Università degli Studi di Bologna, Campus di Cesena, Dipartimento di Informatica - Scienza e Ingegneria, Italia, con una duración global de 10 meses.

5. Enfoque metodológico y etapas de desarrollo

La figura 5.1 ilustra a rasgos generales el enfoque metodológico que guiará las actividades de desarrollo de software / investigación que se llevarán a cabo durante el período sabático, las cuales se listan a continuación. Las etapas de desarrollo

que involucran las actividades antes relacionadas pueden apreciarse en el diagrama de Gantt que se ilustra en la figura 5.2.

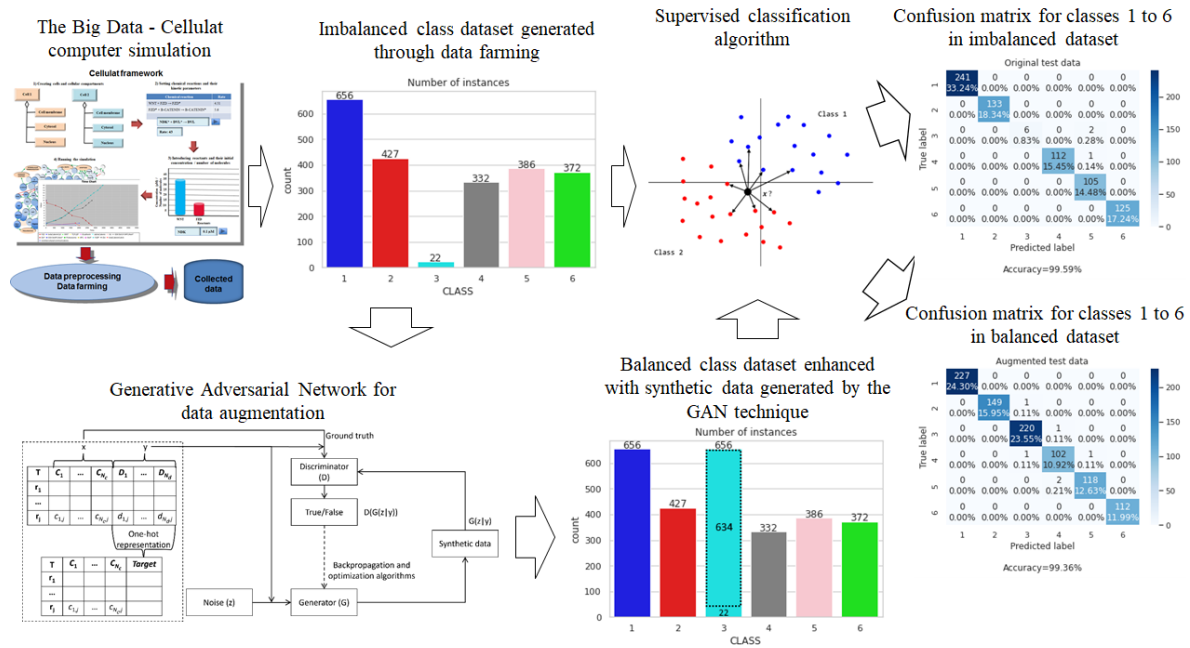


Figura 5.1. Enfoque metodológico que guiará las actividades a desarrollar en el período sabático.

1. Ingeniería inversa e ingeniería progresiva de la plataforma bioinformática Cellulat, con el objetivo de incluir la nueva funcionalidad “generación de datos a gran escala”.
2. Conclusión del desarrollo (ingeniería progresiva) de la herramienta de minería de datos *IDA WEB TOOL*, así como la fase de pruebas.
3. Modelado, simulación y experimentación in silico de vías de señalización celular relacionadas con cáncer en la nueva versión de la plataforma bioinformática, ahora nombrada *Big Data – Cellulat*, y la consecuente generación de grandes volúmenes de datos de calidad.
4. Preparación y exploración (minería de datos) del conjunto de datos a gran escala generado por la plataforma bioinformática *Big Data – Cellulat*, utilizando la herramienta de minería de datos *IDA WEB TOOL*.
5. Identificación del problema del desbalance de clases en el conjunto de datos original producido por la plataforma bioinformática *Big Data – Cellulat*.
6. Aumento del conjunto de datos original producido por la plataforma bioinformática *Big Data – Cellulat*, con datos sintéticos generados por una red generativa-adversativa (GAN).

7. Aplicación del algoritmo de aprendizaje automatizado supervisado al conjunto de datos original producido por la plataforma bioinformática *Big Data – Cellulat*. Análisis de los resultados.
8. Aplicación del algoritmo de aprendizaje automatizado supervisado al conjunto de datos aumentado con la técnica GAN. Análisis de los resultados.

Actividad	Oct 2025	Nov 2025	Dic 2025	Ene 2026	Feb 2026	Mar 2026	Abr 2026	May 2026	Jun 2026	Jul 2026	Ago 2026	Sep 2026	Oct 2026	Nov 2026
Ingeniería inversa e ingeniería progresiva de la plataforma bioinformática <i>Cellulat</i>														
Conclusión del desarrollo (ingeniería progresiva) de la herramienta de minería de datos <i>IDA WEB TOOL</i>														
Experimentación in silico en la plataforma <i>Big Data - Cellulat</i> y generación de datos a gran escala														
Fase de minería de datos sobre los conjuntos de datos generados por la plataforma <i>Big Data - Cellulat</i>														
Enfoque metodológico para abordar el problema del desbalance de clases en conjuntos de datos a gran escala														
Escritura del libro "Biología Computacional y Bioinformática: Un Enfoque práctico"														
Escritura del primer artículo de investigación														
Escritura del segundo artículo de investigación														
Escritura del tercer artículo de investigación														
Presentación de seminarios de investigación y trabajos en eventos académicos*														

*: los meses en los que se indica la presentación de seminarios de investigación y trabajos en eventos académicos son tentativos

Figura 5.2. Diagrama de Gantt que refleja las fases y actividades a desarrollar durante el periodo sabático.

6. Vinculación con los proyectos de investigación, planes y programas académicos de la UAM, Unidad Cuajimalpa

La tabla 6.1 relaciona las actividades a desarrollar durante el periodo sabático y su relación con proyectos de investigación y planes y programas de estudio de la UAM, Unidad Cuajimalpa.

Tabla 6.1: Vinculación de las actividades a desarrollar con los proyectos de investigación, planes y programas de estudio de la UAM, Unidad Cuajimalpa

Actividad a desarrollar durante el periodo sabático	Proyectos de investigación, planes y programas de estudio de la UAM que favorece
<p>Ingeniería de software:</p> <ul style="list-style-type: none"> ❖ Ingeniería inversa e ingeniería progresiva de la Plataforma bioinformática <i>Cellulat</i> ❖ Ingeniería progresiva de la Herramienta de minería de datos <i>IDA WEB TOOL</i> 	<ul style="list-style-type: none"> ▪ Proyecto de investigación divisional “Metodologías para la creación de Sistemas Computacionales y sus aplicaciones en Teoría de Gráficas” ▪ Área temática “Sistemas Computacionales y Bioinformática” del Posgrado en Ciencias Naturales e Ingeniería (PCNI). ▪ Programas de Estudio de la Licenciatura en Ingeniería en Computación: <ul style="list-style-type: none"> ○ UEA Fundamentos de Ingeniería de Software ○ UEA Desarrollo de Software a Gran Escala ○ UEA Proyecto de Ingeniería de Software I ○ UEA Proyecto de Ingeniería de Software II ▪ Programas de Estudio del Posgrado en Ciencias Naturales e Ingeniería (PCNI): <ul style="list-style-type: none"> ○ UEA Bioinformática I ○ UEA Bioinformática II ○ UEA Sistemas Computacionales
<p>Libros de texto:</p> <ul style="list-style-type: none"> ❖ “Biología Computacional y Bioinformática: Un enfoque Práctico” 	<ul style="list-style-type: none"> ▪ Proyecto de investigación divisional “Metodologías para la creación de Sistemas Computacionales y sus aplicaciones en Teoría de Gráficas” ▪ Área temática “Sistemas Computacionales y Bioinformática” del Posgrado en Ciencias Naturales e Ingeniería (PCNI).

	<ul style="list-style-type: none"> ▪ Programas de Estudio del Posgrado en Ciencias Naturales e Ingeniería (PCNI): <ul style="list-style-type: none"> ○ UEA Bioinformática I ○ UEA Bioinformática II ○ UEA Sistemas Computacionales
Artículos de investigación: <ul style="list-style-type: none"> ❖ “Dealing with Medical Imbalanced Datasets in Supervised Machine Learning: An Approach Based on Augmentation and Interpolation Techniques” ❖ “Increasing computer simulation in molecular biology with data farming and synthetic data: a case study in cell signaling networks” ❖ Artículo relacionado con el modelado y simulación de las vías de señalización asociadas a atrofia e hipertrofia muscular 	<ul style="list-style-type: none"> ▪ Proyecto de investigación divisional “Metodologías para la creación de Sistemas Computacionales y sus aplicaciones en Teoría de Gráficas” ▪ Área temática “Sistemas Computacionales y Bioinformática” del Posgrado en Ciencias Naturales e Ingeniería (PCNI).

7. Referencias

- Abufadda, M., & Mansour, K. (2021). A Survey of Synthetic Data Generation for Machine Learning. *2021 22nd International Arab Conference on Information Technology (ACIT)*, 1-7. <https://doi.org/10.1109/ACIT53391.2021.9677302>
- AGUILAR, L. J. (2013). *BIG DATA: ANALISIS DE GRANDES VOLUMENES DE DATOS EN ORGANIZACIONES*. Alfaomega.
- Alon, U. (2019). *An Introduction to Systems Biology*.
- Baxevanis, A. D., Bader, G. D., & Wishart, D. S. (Eds.). (2020). *Bioinformatics*.
- Cárdenas-García, M., González-Pérez, P. P., Montagna, S., Cortés, O. S., & Caballero, E. H. (2016). Modeling Intercellular Communication as a Survival Strategy of Cancer Cells: An In Silico Approach on a Flexible Bioinformatics Framework. *Bioinformatics and Biology Insights*, 10, 5-18. <https://doi.org/10.4137/BBI.S38075>
- Cardenas-Garcia, M., & Pérez, P. (2018). *An in Silico Approach for Understanding the Complex Intercellular Interaction Patterns in Cancer Cells* (p. 195). <https://doi.org/10.5220/0006722601880195>
- Compeau, P., & Pevzner, P. (2018). *BIOINFORMATICS ALGORITHMS*.
- Ewens, W. J., & Grant, G. R. (2004). *Statistical Methods in Bioinformatics: An Introduction*.
- González Pérez, P. P., Omicini, A., & Sbaraglia, M. (2013). A biochemically inspired coordination-based model for simulating intracellular signalling pathways. *Journal of Simulation*, 7(3), 216-226. <https://doi.org/10.1057/jos.2012.28>

- Goyal, M., & Mahmoud, Q. H. (2024). A Systematic Review of Synthetic Data Generation Techniques Using Generative AI. *Electronics*, 13(17), Article 17. <https://doi.org/10.3390/electronics13173509>
- Hernandez, M., Epelde, G., Alberdi, A., Cilla, R., & Rankin, D. (2022). Synthetic data generation for tabular health records: A systematic review. *Neurocomputing*, 493, 28-45. <https://doi.org/10.1016/j.neucom.2022.04.053>
- Jabalia, D. N., & Lakshmi, N. J. (2020). *Bioinformatics, System Biology and Big Data Analysis: Emerging Trends: Bioinformatics, System Biology and Big Data Analysis*.
- Kelley, S. T., & Didulo, D. (2018). *Computational Biology: A Hypertextbook*.
- Lechler, T., Sjarov, M., & Franke, J. (2021). Data Farming in Production Systems—A Review on Potentials, Challenges and Exemplary Applications. *Procedia CIRP*, 96, 230-235. <https://doi.org/10.1016/j.procir.2021.01.156>
- Lesk, A. (2019). *Introduction to Bioinformatics*.
- Marr, B. (2015). *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*. Wiley.
- Marr, B. (2016). *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. Wiley.
- Marz, N., & Warren, J. (2015). *Big Data: Principles and best practices of scalable realtime data systems*. Manning Publications.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: La revolución de los datos masivos*. Turner Noema.
- Mayer-Schönberger, V., & Cukier, K. (2017). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. John Murray.
- Murtaza, H., Ahmed, M., Khan, N. F., Murtaza, G., Zafar, S., & Bano, A. (2023). Synthetic data generation: State of the art in health care domain. *Computer Science Review*, 48, 100546. <https://doi.org/10.1016/j.cosrev.2023.100546>
- Ohyanagi, H., Yamamoto, E., Kitazumi, A., & Yano, K. (2022). *Plant Omics: Advances in Big Data Biology*.
- O'Neil, S. T. (2017). *A Primer for Computational Biology*.
- Ortiz-González, A., González-Pérez, P. P., Cárdenas-García, M., & Hernández-Linares, M. G. (2022). In silico Prediction on the PI3K/AKT/mTOR Pathway of the Antiproliferative Effect of O. joconostle in Breast Cancer Models. *Cancer Informatics*, 21, 11769351221087028. <https://doi.org/10.1177/11769351221087028>
- Sánchez-Gutiérrez, M. E., & González-Pérez, P. P. (2022). Modeling and Simulation of Cell Signaling Networks for Subsequent Analytics Processes Using Big Data and Machine Learning. *Bioinformatics and Biology Insights*, 16, 11779322221091739. <https://doi.org/10.1177/11779322221091739>
- Tolk, A. (2015). The next generation of modeling & simulation: Integrating big data and deep learning. *Proceedings of the Conference on Summer Computer Simulation*, 1-8.
- Waterman, M. S. (1995). *Introduction to Computational Biology: Maps, Sequences, and Genomes*.
- Whitlock, M. C., & Schluter, D. (2014). *The Analysis of Biological Data, Second Edition*.

Ye, S. Q. (Ed.). (2021). *Big Data Analysis for Bioinformatics and Biomedical Discoveries*.
Zvelebil, M., & Baum, J. (2007). *Understanding Bioinformatics*.

A handwritten signature in blue ink, consisting of a large, stylized 'P' followed by a series of loops and a final flourish.

Elaborado por:
Dr. Pedro Pablo González Pérez
15 de marzo de 2025