



**GOBIERNO DE  
MÉXICO**

**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



**TECNOLÓGICO  
NACIONAL DE MÉXICO**

**cenidet**  
Centro Nacional de Investigación  
y Desarrollo Tecnológico

# **CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO**

**Doctorado en Ciencias en Ingeniería Electrónica**  
Especialidad en Control Automático

**Reporte técnico de final de semestre**  
3° semestre

**Modelado de datos ómicos con técnicas de aprendizaje  
profundo mediante redes neuronales**

Presenta:  
**M.C. Arturo Figueroa Arcos**

**Directores de tesis:**

Dr. Víctor Manuel Alvarado Martínez Dra.  
Ma. Guadalupe López López

**Revisores de tesis:**

Dr. Marisol Cervantes Bobadilla  
Dr. Victor Hugo Olivares Peregrino  
Dra. Mónica Borunda Pacheco

Cuernavaca, Morelos a 9 de junio de 2025

# Tabla de contenido

<b>Índice de figuras</b>	<b>ii</b>
<b>Índice de tablas</b>	<b>iii</b>
<b>Nomenclatura</b>	<b>iv</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Objetivos</b>	<b>2</b>
2.1. Objetivo general . . . . .	2
2.2. Objetivos específicos . . . . .	2
<b>3. Marco Teorico</b>	<b>2</b>
3.1. Redes Neuronales Graficas . . . . .	2
3.1.1. Grafos . . . . .	3
3.1.2. GAT . . . . .	3
3.1.3. GCN . . . . .	3
3.1.4. GTN . . . . .	3
<b>4. Avances</b>	<b>3</b>
4.1. Procesamiento de datos ómicos . . . . .	3
4.2. Estructuras DL en datos ómicos . . . . .	3
<b>5. Actividades futuras</b>	<b>3</b>
<b>6. Cronograma de actividades</b>	<b>3</b>

## Índice de figuras

3.1. Ilustración de grafo de entrada (a) y el proceso de la GNN en el cual se realiza el cálculo de la representación vectorial del nodo E agregando información de los nodos vecinos (b). . . . .	3
6.1. Cronograma de actividades . . . . .	4

# Índice de tablas

0.1. Siglas y acrónimos . . . . .	iv
-----------------------------------	----

# Nomenclatura

Tabla 0.1: Siglas y acrónimos

Siglas	Descripción
IA	Inteligencia Artificial
DL	Aprendizaje Profundo (En inglés Deep Learning)
ML	Aprendizaje Automático (En inglés Machine Learning)
CNN	Red Neuronal Convolucional (En inglés Convolutional Neuronal Network)
RNN	Red Neuronal Recurrente (En inglés Recurrent Neuronal Network)
NGS	Secuenciación de nueva generación (En inglés Next Generation Sequencing)
ANN	Red Neuronal Artificial (En inglés Artificial Neuronal Network)
TCGA	Programa del Atlas del Genoma del Cáncer (En inglés The Cancer Genome Atlas Program)
AE	Auto Codificador (En inglés Auto Encoder)
GAN	Red Generativa Antagónica (En inglés Generative Adversarial Network)
DNN	Red Neuronal Densa (En inglés Deep Neural Network)
MLP	Perceptrón Multicapa (En inglés Multilayer Perceptron)

# 1. Introducción

En la ingeniería biomédica se aplica tecnología de última generación, para la creación de dispositivos médicos y métodos que permitan contribuir al bienestar humano, para obtener una mejora en la comprensión de los procesos biológicos que suceden en el ser humano. En este campo de estudio intervienen áreas como la ingeniería electrónica, la ingeniería mecánica, la medicina, la biología, la física, entre otros, considerando a la ingeniería biomédica como un campo interdisciplinario.

La generación de datos ómicos ha experimentado un crecimiento exponencial gracias a las tecnologías de secuenciación de alto rendimiento (NGS). Debido a la alta cantidad de datos existentes, se han impulsado métodos analíticos avanzados para interpretarle y extraer información biológica significativa. Los datos que se abarcan son de diferentes niveles de organización molecular, como el ADN, el ARN, las proteínas y metabolitos.

Algunos ejemplos de estos datos ómicos son los genómicos que estudian el ADN incluyendo la secuencia, estructura y función, los datos transcriptómicos que estudian el ARN incluyendo su expresión, splicing y modificaciones, los datos proteómicos que estudian las proteínas, incluyendo su estructura, función e interacciones y los datos metabolómicos que estudian las moléculas pequeñas que interactúan en las reacciones químicas de la célula.

El análisis de este tipo de datos cuenta con el potencial para mejorar significativamente nuestra comprensión de la salud y las enfermedades en un ser humano. Los pasos típicos son el preprocesamiento donde se realiza la limpieza y normalización de los datos con el fin de eliminar errores, control de calidad para identificar y eliminar posibles valores atípicos, análisis estadístico para identificar biomarcadores relevantes así como también modelos predictivos y de asociación y por último la interpretación de los resultados con sentido biológico. Algunos de los desafíos que se presentan en el análisis son el alto volumen de datos, complejidad de los datos y la integración de tipo de datos, es por ello por lo que se considera un desafío computacional.

El Aprendizaje Profundo (DL) es un campo de la inteligencia artificial, que ha demostrado ser una herramienta poderosa para el análisis de datos complejos, está basado en la utilización de redes neuronales artificiales (ANN's) con el fin de identificar patrones complejos, realizar análisis no lineales y modelar relaciones a partir de diferentes tipos de datos ómicos.

Las ANN's y DL aplicadas en la ingeniería biomédica representan una oportunidad para los profesionales de la salud, ya que permiten realizar análisis más rápidos de grandes conjuntos de datos e información médica relevante, mejoras en los métodos de diagnóstico y pronóstico de enfermedades, diseño de terapias personalizadas y mejoras para el bienestar humano. Los principales desafíos que se presentan son la necesidad de grandes conjuntos de datos para entrenamiento de las redes, la interpretabilidad de las redes neuronales para comprender su toma de decisiones y el requerimiento computacional para entrenamiento de las redes (Sarmiento-Ramos, 2020).

Este proyecto se enfocará en el diagnóstico y pronóstico utilizando datos ómicos centrados en el cáncer, estos son los datos genómicos, transcriptómicos y proteómicos. El área de oportunidad que se encuentra es en el preprocesamiento de datos y la codificación debido a la heterogeneidad de los

datos procedentes de distintas bases que se encuentran de libre acceso. El propósito de este estudio estará dirigido en el estudio de algoritmos de aprendizaje profundo e implementarlos y proponer un modelo basado en algoritmos CNN y RNN así como sus variantes, con enfoque de aplicación en el área biomédica con el fin de dar soporte a médicos especialistas en la precisión del diagnóstico y pronóstico del cáncer.

## **2. Objetivos**

### **2.1. Objetivo general**

- Modelar datos genómicos, transcriptómicos y proteómicos utilizando técnicas de aprendizaje profundo, empleando redes neuronales de diseño propio, y configurar la salida del modelo para propósitos de clasificación y predicción, enfocados en el diagnóstico y pronóstico del cáncer.

### **2.2. Objetivos específicos**

- Codificar datos genómicos, transcriptómicos y proteómicos para ser alimentados en las redes neuronales.
- Implementar redes neuronales convolucionada (CNN) y recurrente (RNN), y entrenarlas.
- Proponer y diseñar una red neuronal propia a partir de las dos anteriores.
- Configurar en cada caso la salida, para propósito de clasificación y predicción.
- Enfocar el modelado de las redes neuronales para fines biomédicos en diagnóstico y pronóstico del cáncer.
- Validar los resultados con las bases de datos e incluyendo opinión de especialistas.

## **3. Marco Teorico**

### **3.1. Redes Neuronales Graficas**

Las Redes Neuronales de grafos (GNN) son modelos de aprendizaje profundo en los cuales se procesan datos en representaciones de grafos o redes. Las GNN permiten aprovechar las relaciones que se codifican en los grafos para extraer características significativas y realizar tareas como la clasificación de nodos, la predicción de aristas (conexiones) y la inferencia a nivel de grafo. Este tipo de modelos aprende de las incrustaciones que capturen la información estructural del grafo como las características que se tengan en los nodos como se puede observar en la figura 3.1

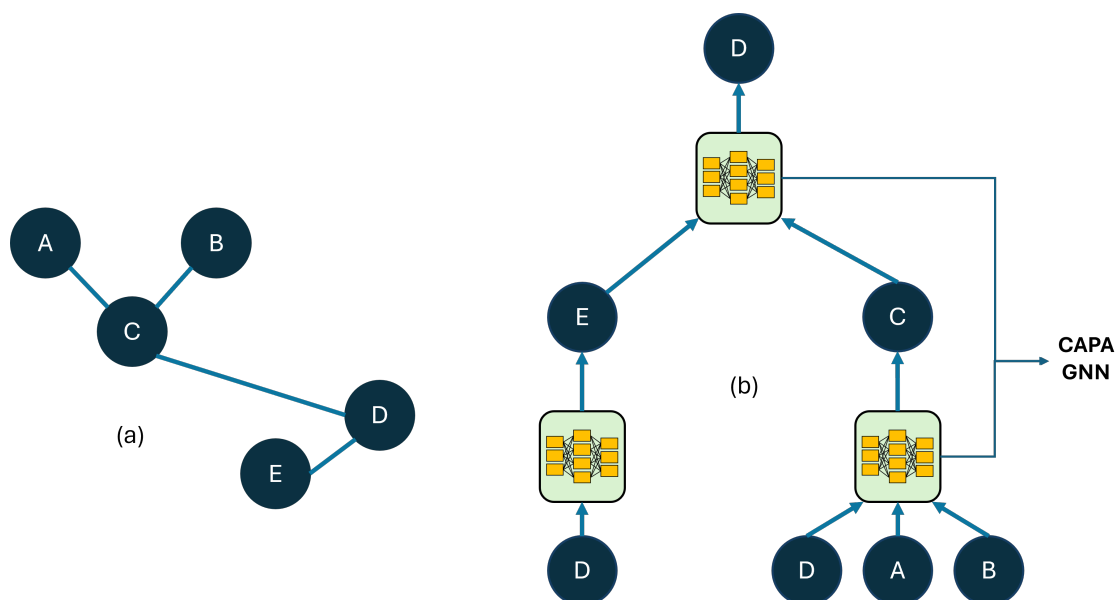


Figura 3.1: Ilustración de grafo de entrada (a) y el proceso de la GNN en el cual se realiza el cálculo de la representación vectorial del nodo E agregando información de los nodos vecinos (b).

### 3.1.1. Grafos

### 3.1.2. GAT

### 3.1.3. GCN

### 3.1.4. GTN

## 4. Avances

### 4.1. Procesamiento de datos ómicos

### 4.2. Estructuras DL en datos ómicos

## 5. Actividades futuras

## 6. Cronograma de actividades

En esta sección se muestra es cronograma de las actividades que de manera preliminar se planean abordar los 4 años de estancia en CENIDET.



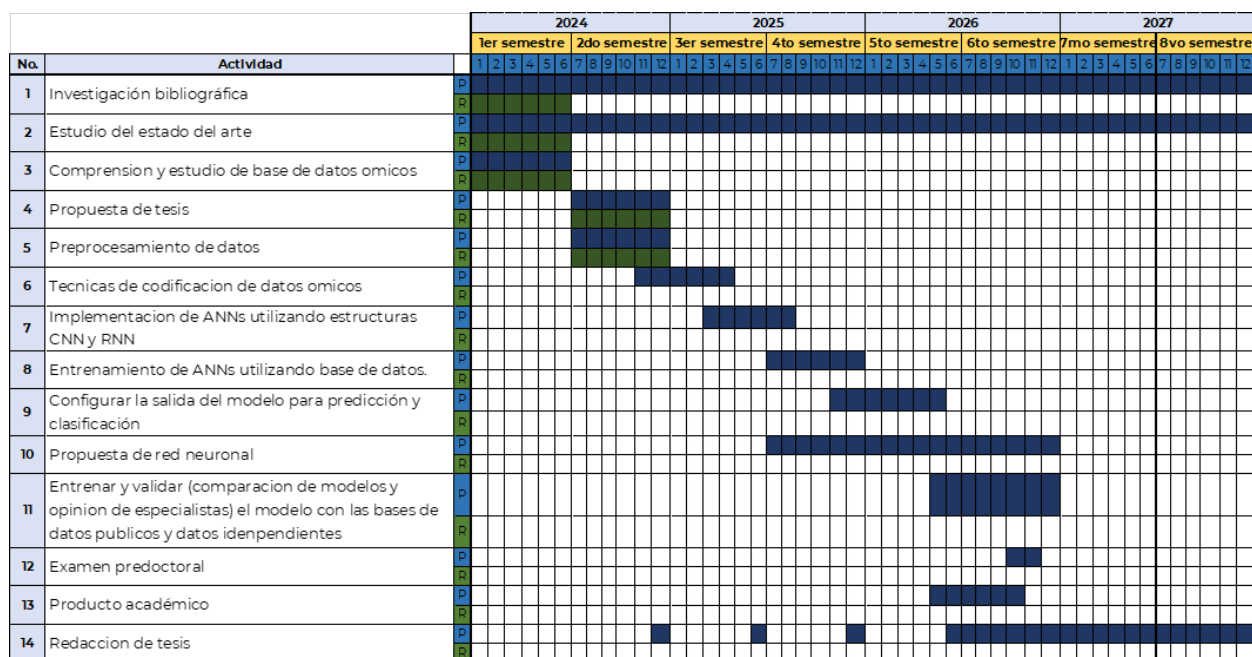


Figura 6.1: Cronograma de actividades

## Referencias

Sarmiento-Ramos, J. L. (2020). Aplicaciones de las redes neuronales y el deep learning a la ingeniería biomédica. *Revista UIS Ingenierías*, 19(4):1–18.