

# Supercómputo

## EN CIMAT

Memorias de proyecto  
“Supercómputo como motor de  
colaboraciones academia - industria”



### Editores

Rafael Herrera Guzmán • Salvador Botello Rionda  
José Jesús Rocha Quezada • Carlos Segura González • Ivete Sánchez-Bravo



Centro de  
Investigación en  
Matemáticas, A.C.



idea  
INDUSTRIA  
INNOVACIÓN  
INTELECTUAL  
INTEGRACIÓN  
INDUSTRIAL

VALLE DE LA  
MENTE  
FACTURA  
GUANAJUATO



CIMNE<sup>®</sup>



Memorias de Proyecto:

Supercómputo como motor de  
colaboración academia - industria



Memorias de Proyecto:

Supercómputo como motor de colaboración  
academia - industria

Editores

Rafael Herrera

Salvador Botello

Jesus Rocha

Carlos Segura

Ivete Sánchez-Bravo

Publicado por:



Centro de  
Investigación en  
Matemáticas, A.C.



**Memorias de Proyecto:  
Supercómputo como motor de colaboración  
academia - industria**

Editores: Rafael Herrera Guzmán, Salvador Botello Rionda, José  
Jesús Rocha Quezada, Carlos Segura González, Ivete Sánchez-Bravo.

ISBN: 978-84-128987-1-2

©Derechos reservados:

Centro de Investigación en Matemáticas A.C.  
Jalisco s/n, Mineral de Valenciana  
Guanajuato, Gto., México. CP 36023  
<https://www.cimat.mx>

Centro Internacional de Métodos Numéricos en la Ingeniería  
Edificio C1, Campus Norte UPC  
Gran Capitán s/n  
Barcelona, España. CP 08034  
<https://www.cimne.com>

Idea GTO  
Edificio Red Estatal Supercómputo  
Mineral de Cata 1305, Puerto Interior  
Silao, Gto., México. CP 36275  
<https://idea.guanajuato.gob.mx>

Primera edición: Septiembre 2024

Este libro no puede ser reproducido total ni parcialmente, por ningún medio electrónico o de otro tipo, sin autorización por escrito de los autores.

This book may not be reproduced, whole or in part, by any means without written permission from the publisher.

Diseño de portada: Idalí Amaya Chagolla

## Prologo:

En la era de la información, la capacidad de procesar y analizar grandes cantidades de datos es fundamental para impulsar el progreso científico, tecnológico, económico y social. La supercomputación ha emergido como una herramienta clave para abordar los desafíos más complejos de nuestra época, desde la simulación de fenómenos climáticos hasta el diseño de nuevos materiales y medicamentos. Actualmente muchas aplicaciones de Inteligencia Artificial requieren de Supercómputo.

Sin embargo, el desarrollo de proyectos de supercomputación es un proceso complejo que requiere la convergencia de múltiples disciplinas, desde la arquitectura de hardware hasta la programación de software, pasando por la gestión de datos y la visualización de resultados.

Para facilitar el acceso de nuevos usuarios al uso del Supercómputo el CIMAT desarrollo del proyecto **Supercómputo como impulsor de colaboraciones academia-industria**, con numero de convenio IDEAGTO/CONV/077/2023 CIMAT, que se llevó a cabo del 31 octubre del 2023 al 31 de octubre de 2024.

En este proyecto, participaron investigadores, estudiantes y empresas interesadas en el uso del Supercómputo como herramienta para resolver sus problemas. En este libro se recopilan y relatan sus experiencias en el uso de los equipos de Supercómputo, a los que tuvieron acceso durante un periodo de 6 meses, que el CIMAT puso a su disposición en Puerto Interior y Guanajuato Capital.

Esperamos que este libro sea de utilidad, a investigadores, ingenieros y estudiantes, al encontrar las experiencias de un público diverso, que se dedicó en 6 meses a resolver proyectos utilizando Supercómputo.

Rafael Herrera

Salvador Botello



## ÍNDICE GENERAL:

<b>APLICACIÓN DE METODOLOGÍAS DE PARALELIZACIÓN DE MEMORIA DISTRIBUIDA PARA LA RESOLUCIÓN DEL PROBLEMA DE EMPLOYEE ROSTERING CON METAHEURÍSTICAS DE TRAYECTORIA.....</b>	<b>1</b>
ANA V. M. C. Y JUAN G. F. A	
<b>MODELADO QSAR DE LA EFICIENCIA EN LA INHIBICIÓN DE LA CORROSIÓN MEDIANTE FÁRMACOS Y FITOQUÍMICOS .....</b>	<b>25</b>
Alan Miralrio, Carlos Beltrán Pérez, Andrés Antonio Abundis Serrano, Mario Villada, Araceli Espinoza Vázquez	
<b>ALMACENAMIENTO DE HIDRÓGENO EN GRAFENO NITROGENADO DECORADOS CON CÚMULOS METÁLICOS..</b>	<b>35</b>
Alan Miralrio, Eduardo Rangel	
<b>MANTENIMIENTO PREDICTIVO IA-OPTIMIZADO PARA INDUSTRIA 4.0 .....</b>	<b>41</b>
Carlos I. Ramírez Bello, Gabriel Ramos Zacarias, Takawira Joseph Mumanga.	
<b>EVALUACIÓN DE ESTRATEGIAS DE INGENIERÍA DE PROMTS PARA MODELO DE LENGUAJE A GRAN ESCALA (LLMS) EN EL PROCESAMIENTO DEL LENGUAJE NATURAL ACADÉMICO.....</b>	<b>53</b>
Guadalupe H. Rangel-Sánchez, Emma Sánchez-Barragán, Liliana Fausto-Castro, Aideé Hernández- López	
<b>CONSTRUCCIÓN DE MODELOS DE LENGUAJE TIPO BERT PARA EL ESPAÑOL MEXICANO .....</b>	<b>90</b>
Luis G. Ruiz Velázquez, Eric S. Téllez Avila, Mario Graff Guerrero, Daniela Moctezuma.	

**OPTIMIZACIÓN GLOBAL DE CLUSTERES DE BORO Y DE PLATINO DOPADOS CON METALES DE TRANSICIÓN ..... 101**  
Peter L. Rodríguez-Kessler, Alvaro Muñoz-Castro

**CLASIFICACIÓN DE ACTIVIDADES HUMANAS ANÓMALAS EN UN ENTORNO DE VIDEO VIGILANCIA ..... 117**  
Jaqueline Reynosa-Guerrero, Daniela Moctezuma, Alberto Vázquez-Cervantes

**IMPACTO DE LAS INTERACCIONES HIDRODINÁMICAS Y ELECTROSTÁTICAS EN LA DINÁMICA DE CONTAGIO DE VIRUS. UNA PERSPECTIVA MESOSCÓPICA..... 130**  
K. Gonzales, S. Ruano, I.A. Segoviano, F. Alarcon



# APLICACIÓN DE METODOLOGÍAS DE PARALELIZACIÓN DE MEMORIA DISTRIBUIDA PARA LA RESOLUCIÓN DEL PROBLEMA DE EMPLOYEE ROSTERING CON METAHEURÍSTICAS DE TRAYECTORIA

**Ana V. M. C. y Juan G. F. A.**

*Optimen*

*Av. Leon #119, 4th floor.*

*Col. Jardines del Moral Leon, Gto. Mexico, 37160.*

*Email: [info@optimen.com.mx](mailto:info@optimen.com.mx).*

*web page: <https://www.optimen.com.mx/>*

**Resumen.** El propósito de este estudio es investigar diversas metodologías de paralelización de memoria distribuida aplicadas a un problema de optimización de asignación de turnos laborales, conocido como Employee Rostering, mediante el uso de metaheurísticas de trayectoria. Este problema implica la asignación de turnos laborales bajo múltiples restricciones y variables, lo que lo convierte en un problema NP-Hard. Se implementaron dos metaheurísticas de trayectoria, Hill Climbing y Simulated Annealing, a las cuales se les aplicó una paralelización funcional de tipo MIR (Multiple Independent Runs), con el objetivo de analizar la eficacia de esta metodología de paralelización al optimizar los costos operativos, al tiempo que se cumplen las restricciones establecidas. Aunque este análisis introductorio destaca las fortalezas y limitaciones de cada método paralelizado, también proporciona una guía para su implementación en algoritmos de optimización más complejos, como los algoritmos genéticos.

**Palabras clave:** Employee Rostering, Optimización, Metaheurísticas de trayectoria, Hill Climbing, Simulated Annealing, Paralelización.

## INTRODUCCIÓN

Los problemas de planificación, como el Employee Rostering, comúnmente implican la asignación de un conjunto de turnos de trabajo a una cantidad determinada de recursos, usualmente referido a empleados, en donde se cuida satisfacer restricciones y minimizar los costos asociados. Este tipo de problemas tiene una amplia variedad de aplicaciones en casi cualquier entorno empresarial que dependa de capital humano, tales como bancos, escuelas, hoteles, hospitales, call centers, manufactura, construcción, agencias

gubernamentales, aerolíneas y empresas de transporte y logística en general. El poder lograr una óptima programación de las tareas del personal tiene como ventaja el generar un porcentaje de ahorro, el cual puede provenir del uso eficiente de la nómina, aumento de ingresos, minimización de tiempos, planificación de movimientos, etc. lo que al final provoca un aumento en la competitividad de la empresa.

Una aplicación de este caso muy estudiado es en el área de la salud, donde la gestión del personal de hospitales y centros de especialidades representa un gran reto, además de una tarea para realizar con cuidado, ya que se sabe que una administración inapropiada de los recursos puede llevar a malas experiencias de los pacientes, insatisfacción del personal, daño a la salud y el bienestar, y en general, un bajo rendimiento del hospital. Otra aplicación conocida es en el sector de atención al cliente de los call centers, en estos casos el objetivo de la optimización suele ser basado en costos operacionales, satisfacción del empleado y tiempos de espera en el servicio al cliente; esto en respuesta a patrones extremadamente dinámicos de demanda, dados en función de meses, semanas, días, e incluso horas, en esquemas 24/7.<sup>1</sup> El problema de Employee Rostering suele consistir en dos etapas, aunque en algunos casos se manejan al mismo tiempo, el primero paso es la generación de turnos que cumplan ciertas características óptimas (también conocido como un proceso de "pairing"), para después puedan ser asignados al personal (este segundo proceso de asignación se conoce como "rostering"). Algunos autores abordan la solución de este último tipo de problemas a través de aproximaciones con programación dinámica<sup>2</sup>.

Debido a que este tipo de problemas suelen ser descritos con una enorme cantidad de restricciones que se deben satisfacer al mismo tiempo, se vuelve realmente complicado poder llegar a una solución óptima. Estas restricciones comúnmente giran en torno a contratos de trabajo, ya sean individuales o colectivos, reglamentaciones gubernamentales y/o sindicales, así como políticas internas de la propia empresa. Otra complejidad proviene desde la ventana temporal sobre la cual se debe *rostar* una asignación, la cual puede variar desde un día o hasta un mes, y durante la cual el empleado puede realizar varias tareas diferentes dentro de un mismo turno. En cuanto a la labor asignada *per se*, pueden existir tareas que son demasiado especializadas y diversificadas y que, por tanto, requieren ciertas calificaciones por parte del empleado para llevarse a cabo, añadiendo aún más restricciones al problema. Por esta conjunción de razones, en sus muchas variantes, el problema de rostering y sus subproblemas relacionados han sido clasificados como NP-Hard y NP-complete y, por tanto, son extremadamente difíciles de resolver de manera óptima<sup>3</sup>.

A pesar de las revisiones exhaustivas sobre métodos y modelos para el problema de Rostering<sup>4</sup>, la variabilidad de los factores que determinan estos problemas hace que sea difícil implementar un software universal. Por esta razón, muchas empresas aún prefieren métodos manuales o semi-manuales para programar a su personal. Se necesitan soluciones de software que frezcan un motor de optimización escalable y potente, así como herramientas versátiles para generar múltiples escenarios y modelar diferentes casos de negocio.

Un método efectivo para poder llegar a soluciones aceptables en un tiempo razonable es gracias a la paralelización de los algoritmos de metaheurísticas. Para el caso de los problemas NP-Hard existen distintas estrategias que se pueden implementar, como es la paralelización a nivel de datos, paralelización a nivel de tareas y paralelización a nivel de soluciones. Las principales ventajas de utilizar este paradigma son evidentes, puesto que si se implementa de manera correcta se puede lograr una gran reducción de tiempos de ejecución, ya que podemos utilizar distintos procesos para dividir tareas pesadas o complejas. Se ha notado también, en la utilización de estos métodos, una mejora en la calidad de las soluciones ya que nos permite explorar mejor el espacio de búsqueda desde diferentes puntos de inicios aleatorios; además de que mejora en la robustez y escalabilidad del algoritmo para adaptarse mejor a cambios en el entorno o en los datos<sup>5</sup>.

## MARCO TEÓRICO

Las metaheurísticas son algoritmos de optimización que buscan soluciones aproximadas a problemas difíciles (frecuentemente NP-hard) mediante la combinación de técnicas de búsqueda exploratorias e intensivas. Los algoritmos de trayectoria, como el Simulated Annealing (SA) y Tabu Search (TS), son ampliamente utilizadas para resolver problemas complejos. Estas técnicas combinan diferentes heurísticas y métodos inspirados en inteligencia artificial, evolución biológica y mecánica estadística. Ejemplos incluyen el uso de SA en el problema de Rostering<sup>6</sup> y TS para considerar preferencias de turnos en enfermería<sup>7</sup>.

Los algoritmos de Simulated Annealing paralelos (PSA) han sido revisados e introducidos ampliamente por distintos autores<sup>8,9,10</sup>, entre estos temas también se ha estudiado la convergencia de varios algoritmos PSA a través de

movimientos paralelos<sup>11</sup>. Sin embargo, hasta el momento no existe una clasificación ampliamente aceptada de PSA<sup>5</sup>. Se han probado distintas maneras de paralelización, de las cuales se destacan dos: una es mediante la partición adecuada del algoritmo<sup>12,13</sup> y la otra por la paralelización de los datos.

La infraestructura de la paralelización se implementa principalmente a partir de tecnologías de máquinas paralelas: SIMD (Single Instruction Multiple Data), MIMD (Multiple Instruction Multiple Data) y redes de computadoras, como LAN (Local Area Network) y WAN (Wide Area Network), para este estudio en particular principalmente se hablará de SIMD, MIMD las cuales están basadas en el uso de múltiples procesadores, que están acoplados de manera suelta o estrecha. Estos recursos se utilizan para procesar las tareas en paralelo para cualquiera de las dos metodologías<sup>14,15,16</sup>. Las comunicaciones necesarias entre las unidades paralelas se manejan utilizando frameworks de Interfaces de Paso de Mensajes Message Passing Interface (MPI), los cuales en este estudio se implementaron en el lenguaje de programación de C++, para ser ejecutados en los servidores del clúster de super cómputo del Centro de Investigación en Matemáticas (CIMAT).

Dentro de la clasificación de la programación distribuida, existen dos formas posibles de paralelismo: (a) el paralelismo físico, que implica la paralelización a través de la descomposición y distribución de los datos o del espacio de búsqueda según los recursos computacionales, y (b) el paralelismo funcional o algorítmico, que considera el algoritmo en sí para este propósito<sup>17</sup>. El paralelismo funcional se aplica, por ejemplo, al Simulated Annealing de dos maneras: (i) lanzando múltiples ejecuciones independientes (MIR) concurrentemente con diferentes condiciones iniciales, y (ii) explorando el estado de equilibrio concurrentemente en cada nivel de temperatura. Aunque en la literatura no se ha estudiado mucho el algoritmo de Hill Climbing en un entorno de paralelización, en este estudio se realizó con la misma metodología MIR para hacer la comparación entre ambos algoritmos.

El paralelismo de Ejecuciones Independientes Múltiples (MIR) es uno de los métodos más populares debido a su facilidad de uso. La idea principal es lanzar múltiples ejecuciones del mismo algoritmo de forma simultánea, cada una ejecutándose en diferentes procesos con soluciones iniciales distintas, al finalizar las ejecuciones, se selecciona el mejor resultado entre todas las salidas [Figura 1]. Este enfoque ha sido utilizado por muchos investigadores y es la forma más sencilla de implementar el paralelismo<sup>5</sup>. La construcción de comunicaciones entre procesos puede ser diseñada como sistema cliente-servidor, donde las tareas administrativas, cómo la asignación de tareas entre procesos es realizada únicamente por la unidad central, mientras que cada

ejecución independiente funciona como un cliente, el cual al final del proceso solo presenta el resultado producido al servidor.

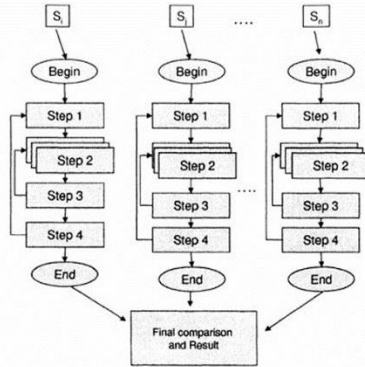


Figura 1. Modelo de paralelización por MIR.

La principal dificultad radica en decidir si se debe o no construir la interacción entre las ejecuciones independientes y cómo gestionar la comunicación. Esto puede afectar la calidad de la solución dependiendo del problema abordado. La comunicación entre los procesadores se realiza mediante sistemas de paso de mensajes (MPS) o compartiendo memoria. Dado que la comunicación puede ser más costosa que la computación, debe establecerse un equilibrio entre ambas, dependiendo del problema y las propiedades de la infraestructura disponible.<sup>5</sup>

## METODOLOGÍA

El presente estudio emplea técnicas de paralelización de memoria distribuida utilizando el servidor de super computo del Centro de Investigación en Matemáticas (CIMAT). El código base del optimizador fue desarrollado en el lenguaje de programación C++, y para la comunicación entre procesos se utilizó la biblioteca Message Passing Interface (MPI).

Inicialmente, se programó un optimizador tomando como referencia el caso de estudio de Shiau<sup>18</sup> para la resolución del problema de Employee Rostering. Durante el desarrollo, se implementaron nuevas reglas de negocio con el objetivo de obtener una mayor similitud entre el caso de estudio y las condiciones reales de una operación de negocios. La metodología permanece abierta a la posibilidad de agregar nuevas reglas o preasignaciones del personal para enriquecer el trabajo.



Se implementaron dos metaheurísticas de trayectoria: Hill Climbing y Simulated Annealing. Estas técnicas fueron elegidas debido a su capacidad para proporcionar buenas soluciones en tiempos de ejecución relativamente cortos. Ambas metaheurísticas parten de una solución inicial y la mejoran a través de una búsqueda local utilizando operadores, lo que las convierte en herramientas robustas para manejar restricciones y objetivos complejos.

A pesar de la eficacia de Hill Climbing y Simulated Annealing, la integración de técnicas de paralelización en la lógica de optimización presentó ciertos desafíos. Se exploraron varias estrategias:

- **Paralelización de la Función de Costo:** Esta estrategia implicaba dividir el total de la asignación entre los nodos para evaluar la solución encontrada. Sin embargo, debido a la baja complejidad del cálculo de la función de costo en este caso de estudio, esta aproximación no resultó viable.
- **División del Espacio de Búsqueda:** Otra aproximación consistía en dividir el espacio de búsqueda para realizar la exploración de la solución. No obstante, esta técnica podría afectar negativamente a los operadores desarrollados, además de impedir la visualización correcta de restricciones de interés, por lo que no se implementó.
- **Paralelización funcional servidor-cliente:** En este enfoque, cada nodo ejecutaba el proceso de optimización de manera independiente y, al finalizar, se seleccionaba la mejor solución encontrada entre todos los nodos. Este método dio buenos resultados al momento de ser implementado.

Para estudiar la mejora obtenida mediante la paralelización, se realizaron pruebas de ejecución en serie y en paralelo. Se evaluaron tanto los tiempos de ejecución como el uso de la memoria. Adicionalmente, se llevaron a cabo pruebas estadísticas para analizar la diferencia de rendimiento entre los dos algoritmos implementados.

## CASO DE ESTUDIO

En este artículo se aborda una versión simplificada del problema de gestión de personal aeroportuario propuesto por Shiau<sup>18</sup>, en el cual se busca asignar a una cantidad de recursos (personal de staff) un horario adecuado para disminuir una demanda mensual de manera que se respeten las restricciones propuestas. Se cuenta con un staff de  $i = \{1, \dots, I\}$  empleados, donde  $I = 35$  en nuestro caso, en donde cada empleado fue asignado con 2 habilidades (spots de trabajo) a los que pueden ser asignados, que pertenecían al conjunto

de  $j = \{1, \dots, J\}$  donde  $J=6$ , para el caso del estudio en un ambiente aeroportuario estas áreas de trabajo podían ser: administración, check-in, despacho, equipaje, embarque, rampa, mantenimiento, etc. El periodo de planeación de manejo de  $d = \{1, \dots, D\}$  con  $D = 28$ , donde cada día consta de  $t = \{1, \dots, T\}$  horas laborales con  $T=24$ , y un conjunto de  $w = \{1, \dots, W\}$  turnos de trabajo. En nuestro caso  $W = 10$  y cada turno tenía como propiedades una hora de inicio, una hora de fin y un total de horas de trabajo asignadas.

La variable de decisión en este caso consistía en el estado de si un turno  $w$  era asignado al empleado  $i$  para el día  $d$ , como se trata de una variable binaria puede ser descrita como:

$$x_{iwd} = \begin{cases} 1, & \text{si fue asignado} \\ 0, & \text{en otro caso} \end{cases} \quad (1)$$

De igual manera, tenemos la variable de decisión  $p_{jwd}$  la cual nos limita a poder asignar al empleado  $i$  en una posición  $j$  a la que este calificado. Su descripción para este caso sería:

$$p_{ijwd} = \begin{cases} 1, & \text{si la posición esta siendo cubierta} \\ 0, & \text{en otro caso} \end{cases} \quad (2)$$

La función objetivo de este caso es para una variable que nos describe el costo de la programación del turno que se quiere minimizar, este se representa por la siguiente expresión en donde  $k$  es el costo por hora trabajada, el cual es constante para todos los empleados, y  $h_w$  corresponde al número de horas trabajadas durante el turno  $w$ .

$$f = \sum_{i=1}^I \sum_{w=1}^W \sum_{d=1}^D k * h_w * x_{iwd} \quad (3)$$

Esta función de costo está sujeta a varias restricciones que se fueron implementando, sin embargo, hasta el momento son pocas a comparación de la complejidad real del problema, se busca ir incorporando nuevas reglas conforme los algoritmos se vayan refinando.

La primera restricción representa el cumplimiento de la demanda de empleados para una determina posición de trabajo  $j$ , un día  $d$  y una hora  $t$ , y está definida por la variable  $r_{jdt}$ . Esta variable fue representada de forma matricial donde se plasma la necesidad de cobertura para cada una de las posiciones a lo largo de toda la jornada diaria a nivel de hora durante un

periodo definido, e.g. una semana:

$$\sum_{i=1}^I \sum_{w=1}^W e_{wj} * p_{ijwd} \geq \forall j \in J, \forall t \in T, \forall d \in D \quad (4)$$

La variable  $e_{wt}$  representa si el turno  $w$  fue trabajado en la hora  $t$ :

$$e_{wt} = \begin{cases} 1, & \text{si se cubre la hora } t \text{ con el turno } w \\ 0, & \text{en otro caso} \end{cases} \quad (5)$$

La segunda restricción nos indica que un empleado solo puede ser asignado a un máximo de un turno por día:

$$\sum_{w=1}^W x_{iwd} \leq 1, \forall i \in I, \forall d \in D \quad (6)$$

La tercera restricción establece que cada empleado solo puede ser asignado a una posición de trabajo en un turno, siempre y cuando se cuente con la cualificación  $q_{ij}$  requerida para tal posición:

$$p_{ijwd} \leq q_{ij} * x_{iwd}, \forall i \in I, \forall j \in J, \forall w \in W, \forall d \in D \quad (7)$$

$$\sum_{j=1}^J p_{ijwd} \leq 1 \quad (8)$$

Las cualificaciones (skills) de cada empleado fueron tomados del artículo de Shiau<sup>18</sup>, cada empleado fue asignado con un total de dos skills [Tabla 1].

Staff ID	j <sub>1</sub>	j <sub>2</sub>	j <sub>3</sub>	j <sub>4</sub>	j <sub>5</sub>	j <sub>6</sub>
1-8	✓	✓				
9-19			✓	✓		
20-35					✓	✓

Tabla 1. Distribución de cualificaciones por rango de Staff ID

La cuarta restricción determina un periodo de descanso reglamentario entre dos turnos consecutivos, es decir, si un empleado trabaja los días  $d$  y  $d+1$ , la variable  $y_{w_1w_2}$  deberá indicar si los turnos  $w_1$  y  $w_2$  puede ser trabajados en días consecutivos. Esta restricción nos permite controlar que un empleado que se

ha desvelado en su turno, el siguiente no tenga que ser en un horario de madrugada, por ejemplo. En este caso la variable  $y_{w_1w_2}$  se describe como:

$$y_{w_1w_2} = \begin{cases} 1, & \text{si el turno } w_2 \text{ puede ser después de } w_1 \\ 0, & \text{en otro caso} \end{cases} \quad (9)$$

Por lo que la condición sería:

$$x_{iw_1d} + x_{iw_2d+1} \leq y_{w_1w_2} + 1, \forall i \in I, \forall w_1, w_2 \in W, \forall d \in D \quad (10)$$

En este estudio se utilizó la secuencia de turnos igual a como la define Shiau<sup>18</sup> [Tabla 2]. Por lo que para poder determinar la legalidad de estos turnos se buscó una representación matricial binaria [Tabla 3]

w	Horarios	Duración
1	5:00-14:00 hrs	9 hrs
2	5:00-15:00 hrs	10 hrs
3	5:00-16:00 hrs	11 hrs
4	5:00-17:00 hrs	12 hrs
5	6:00-17:00 hrs	11 hrs
6	8:00-17:00 hrs	9 hrs
7	13:00-20:00 hrs	7 hrs
8	13:00-23:00 hrs	10 hrs
9	13:00-24:00 hrs	11 hrs
10	14:00-24:00 hrs	10 hrs

*Tabla 2: Conjunto de turnos y horas laboradas*

	1	2	3	4	5	6	7	8	9	10
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
7	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓
8	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓
9	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓
10	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓

*Tabla 3: Restricciones de asignación de turnos en días consecutivos.*

Una quinta restricción determinaría una cota superior e inferior a la cantidad de horas que puede trabajar un empleado en un periodo de una semana laboral:

$$d_c = D \text{ mód } 7 \quad (11)$$

$$\sum_d^{d_c} \sum_t^T x_{iwd} \quad (12)$$

En este caso la primera suma itera por cada día de la semana  $d$ , la segunda suma itera por todas las horas que abarca el turno  $w$  asignado. En nuestro caso dejamos un límite superior de 56 horas y un inferior de 40.

$$40 \leq \sum_{d=d_c}^{d_c+6} \sum_{w=1}^W h_w * x_{iwd} \leq 56, \forall i \in I \quad (13)$$

La sexta restricción limita a un total de 6 días consecutivos de trabajo sin descanso intermedio. Donde  $x_{iwd'}$  es la variable binaria que indica si el empleado  $i$  trabaja en el turno  $w$  en el día  $d'$ . Y  $y_d$  es una variable binaria auxiliar que se activa cuando el empleado  $i$  comienza un nuevo ciclo de trabajo consecutivo en el día  $d$ .

$$\sum_{d'=d}^{d+5} \sum_{w=1}^W x_{iwd'} \leq 6 * y_d, \forall i \in I, \forall d \in \{1, \dots, D - 6\} \quad (14)$$

La séptima restricción nos habla sobre el número máximo y mínimo de días de trabajo que puede realizar un empleado en un periodo de un mes laboral:

$$\sum_d^{D=28} x_{iwd} \quad (15)$$

Para este caso se consideró una asignación de 28 días, la suma itera por todos los días que se les fue asignados trabajar, poniendo como límite 21 días máximos de trabajo y 20 mínimos de trabajo.

La octava restricción establece un número máximo de empleados concurrentes en un tiempo determinado. En este caso se recorre la suma por

cada día de la asignación y se suma el total de empleados asignados para esa fecha, el límite máximo se estableció en 30.

$$\sum_i^I x_{id}, \forall d \in D \quad (16)$$

Los costos de penalización que se utilizaron para cada restricción están mostrados en la tabla [Tabla 4]

Descripción	Costo
Costo por hora por empleado	1
Restricción de turnos ilegales	1000
Demanda baja en la asignación	1000
Demanda alta en la asignación	1
Restricción máx. días consecutivos de trabajo	1000
Restricción máx. horas de trabajo por semana	1000
Restricción mín. horas de trabajo por semana	1000
Restricción máx. días de trabajo por mes	1000
Restricción mín. días de trabajo por mes	1000
Restricción máx. empleados en un día	1000

*Tabla 4: Tabla de los costos asignados para el estudio.*

La demanda de entrada para este caso tiene la siguiente configuración [Tabla 5]

Sunday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
j=1	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	2.5	2.5	1.5	1.5	1.5	1	1	1	1	1	1	1
j=2	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	2.5	3.5	3.5	2.5	2.5	2	2	2	2	2	2	1
j=3	0	0	0	0	2	2	2	3	3	3	3	3	4	4	3	3	2	1	1	1	0	0	0	0
j=4	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0
j=5	0	0	0	0	1	2	2	2	2	2	2	2	2.5	2.5	2.5	1.5	1.5	0.5	0.5	0.5	0	0	0	0
j=6	0	0	0	0	5	5	5	5	5	5	5	5	5.5	6.5	6.5	6.5	6.5	1.5	1.5	1.5	1	1	1	1
Monday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
j=1	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	1.5	2.5	1.5	1.5	1.5	1	1	1	1	1	1	1
j=2	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	2.5	3.5	3.5	2.5	2.5	2	2	2	2	2	2	1
j=3	0	0	0	0	2	2	2	3	3	3	3	3	4	4	4	3	1	1	1	0	0	0	0	0
j=4	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0
j=5	0	0	0	0	1	2	2	3	3	3	3	3	4	4	4	3	3	1	1	1	0	0	0	0
j=6	0	0	0	0	4	5	5	5	5	5	5	5	5	6	6	6	4	2	1	1	1	1	1	1
Tuesday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
j=1	0	0	0	0	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
j=2	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	2.5	2	2	2	2	2	2	2	2	2	2	1
j=3	0	0	0	0	2	2	2	3	3	3	3	3	4	4	4	4	2	1	1	1	0	0	0	0
j=4	0	0	0	0	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0
j=5	0	0	0	0	1	2	2	3	3	3	3	3	4.5	4.5	4.5	4.5	4.5	1.5	1.5	1.5	0.5	0.5	0.5	0.5
j=6	0	0	0	0	2	4	4	4	4	4	4	4	5.5	5.5	5.5	3.5	3.5	1.5	1.5	1.5	0.5	0.5	0.5	0.5
Wednesday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
j=1	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1	1	1	1	1	1	1
j=2	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	2.5	3.5	3.5	2.5	2.5	2	2	2	2	2	2	1
j=3	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	2.5	3.5	3.5	2.5	2.5	2	2	2	2	2	2	1
j=4	0	0	0	0	2	2	2	2.5	2.5	2.5	2.5	2.5	2.5	3.5	3.5	2.5	2.5	1	1	1	1	1	1	0
j=5	0	0	0	0	2	2	2	3	3	3	3	3	4.5	4.5	4.5	1.5	1.5	1.5	1.5	1.5	1	1	1	0
j=6	0	0	0	0	4	4	4	5	5	5	5	5	6.5	6.5	6.5	3.5	3.5	1.5	1.5	1.5	1	1	1	0
Thursday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
j=1	0	0	0	0	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1
j=2	0	0	0	0	1	1	1	1	1	1	1	1	2	3	3	3	2	2	2	2	2	2	2	1
j=3	0	0	0	0	2	2	2	3	3	3	3	3	4	4	4	4	2	1	1	1	1	1	1	0
j=4	0	0	0	0	2	2	2	2	2	2	2	2	3	3	2	2	2	1	1	1	1	1	1	0
j=5	0	0	0	0	2	2	2	2	2	2	2	2	3	4	4	3	2	2	2	2	2	1	1	1
j=6	0	0	0	0	2	3	3	4	4	4	4	4	5	5	5	2	2	1	1	1	1	1	1	0
Friday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
j=1	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	2.5	2.5	2.5	1.5	1.5	1	1	1	1	1	1	1
j=2	0	0	0	0	1	1	1	1.5	1.5	1.5	1.5	1.5	3.5	3.5	3.5	2.5	2.5	2	2	2	2	2	2	1
j=3	0	0	0	0	2	2	2	3.5	3.5	3.5	3.5	3.5	4.5	4.5	4.5	2.5	2.5	1	1	1	1	1	1	0
j=4	0	0	0	0	2	2	2	2.5	2.5	2.5	2.5	2.5	3.5	3.5	2.5	2.5	2.5	1	1	1	1	1	1	0
j=5	0	0	0	0	1	2	2	3	3	3	3	3	4.5	4.5	3.5	3.5	3.5	1.5	1.5	1.5	1	1	1	0
j=6	0	0	0	0	4	4	4	5	5	5	5	5	6.5	6.5	6.5	3.5	3.5	1.5	1.5	1.5	1	1	1	0
Saturday	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
j=1	0	0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	1	1	1	1	1	1	1	1
j=2	0	0	0	0	1	1	1	1	1	1	1	1	2	3	3	3	2	2	2	2	2	2	2	1
j=3	0	0	0	0	2	2	2	3	3	3	3	3	4	4	4	4	2	1	1	1	0	0	0	0
j=4	0	0	0	0	2	2	2	2	2	2	2	2	3	3	2	2	2	1	1	1	1	1	1	0
j=5	0	0	0	0	1	2	2	2	2	2	2	2	3	3	3	3	3	1	1	1	0	0	0	0
j=6	0	0	0	0	2	2	2	2	2	2	2	2	6	6	6	6	4	4	4	4	0	0	0	0

## ALGORITMOS PROPUESTOS

Para este estudio se ha optado por una prueba preliminar con metaheurísticas de trayectoria: Hill Climbing y Simulated Annealing, debido a su facilidad de implementación y que se ha comprobado su eficiencia para resolver problemas de Scheduling. Esto nos sirvió para ir planteando una prueba de paralelización, comprendiendo primeramente en que sección del algoritmo se consumía el mayor tiempo de ejecución y, por lo tanto, el área de interés a paralelizar. Para siguientes trabajos se buscaría la implementación de métodos **poblacionales** haciendo uso del poder computacional y de memoria que se ofrece en espacios como clúster de super computo. Lo que nos permitiría solventar la dificultad de lidiar con problemas a gran escala y poder generar poblaciones grandes de individuos que enriquezcan las soluciones.

El primer algoritmo, Hill Climbing [Figura 2], fue utilizado con el fin de poder conocer el **espacio de búsqueda**, es decir, la región de soluciones factibles de nuestro problema. La mayor desventaja de este algoritmo es la facilidad con la que queda estancado en mínimo o máximos locales, por lo que también se implementó el algoritmo de Simulated Annealing [Figura 3].

El funcionamiento básico del algoritmo de Hill Climbing es el siguiente:

1. *Estado Inicial*: Se comienza con una solución inicial, que puede ser generada aleatoriamente o mediante algún método heurístico. Esta solución inicial es una asignación de turnos a empleados que cumple, en mayor o menor medida, con las restricciones del problema. Inicialmente se tomó como entrada la solución propuesta por Shiau<sup>18</sup>
2. *Evaluación*: Se evalúa la solución inicial utilizando una función objetivo, en este caso se contempla el costo real de la asignación de un empleado en un puesto de trabajo junto el cumplimiento de las restricciones. En este caso podría ser una función de costo que contemple factores como la satisfacción de restricciones y la minimización de costos.
3. *Generación de Vecinos*: A partir de la solución actual, se generan nuevas soluciones, conocidas como vecinos. Los vecinos se obtienen realizando pequeñas modificaciones a la solución actual, para los cuales creamos una serie de operadores que realizan tareas como cambios de turnos.
4. *Selección del Mejor Vecino*: Se evalúan las soluciones vecinas y se selecciona la mejor según la función objetivo. Si esta nueva solución es mejor que la solución actual, se reemplaza la solución actual con esta nueva solución.
5. *Iteración*: Este proceso se repite iterativamente, desde la generación de vecinos, evaluación de la solución generada y la selección del mejor.
6. *Condición de Parada*: El algoritmo continúa iterando hasta que se cumple alguna condición de parada, en este caso se utilizó un número límite de iteraciones.



**Algorithm 1** Hill Climbing

---

```

1: procedure HILLCLIMBING
2:   currentSolution ← solución inicial arbitraria o heurística
3:   currentValue ← Evaluate(currentSolution)
4:   loop
5:     neighbor ← un vecino de currentSolution con mejor
      evaluación
6:     if no hay neighbor o neighbor no mejora a
      currentSolution then
7:       break
8:     end if
9:     currentSolution ← neighbor
10:    currentValue ← Evaluate(currentSolution)
11:  end loop
12:  return currentSolution
13: end procedure

```

---

Figura 2. Pseudo código del algoritmo de Hill Climbing

El funcionamiento del Simulated Annealing se describe a continuación:

1. *Estado Inicial*
2. *Temperatura Inicial*: Se establece una temperatura inicial alta  $T$ , que controlará la probabilidad de aceptar soluciones peores al inicio del proceso. La temperatura se reduce gradualmente a medida que avanza el algoritmo.
3. *Evaluación*
4. *Generación de vecinos*
5. *Aceptación de vecinos*: La nueva solución vecina se evalúa y se decide si se acepta o no, utilizando un criterio probabilístico  $e^{\Delta E/T}$ . Si la nueva solución es mejor que la actual, se acepta automáticamente. Si es peor, se acepta con una probabilidad que depende de la diferencia en la función objetivo y de la temperatura actual, permitiendo así explorar soluciones subóptimas al inicio para escapar de óptimos locales.
6. *Enfriamiento*: La temperatura se reduce según un factor de enfriamiento predefinido  $alpha$ . La reducción de la temperatura disminuye la probabilidad de aceptar soluciones peores a medida que avanza el algoritmo.
7. *Iteración*: El proceso se repite iterativamente, generando y evaluando nuevos vecinos, y actualizando la solución actual según el criterio de aceptación y la temperatura.
8. *Condición de parada*: El algoritmo continúa iterando hasta que se cumple alguna condición de parada, en este caso se estableció una temperatura mínima alcanzable  $T_{min}$ .

---

**Algorithm 2** Simulated Annealing (Recocido Simulado)

---

```
1: procedure SIMULATEDANNEALING
2:   solution  $\leftarrow$  solución inicial
3:   T  $\leftarrow$  temperatura inicial
4:   Tmin  $\leftarrow$  temperatura mínima
5:   alpha  $\leftarrow$  factor de enfriamiento
6:   while T > Tmin do
7:     newSolution  $\leftarrow$  vecino aleatorio de solution
8:     deltaE  $\leftarrow$  costo(newSolution) - costo(solution)
9:     if deltaE < 0 then
10:      solution  $\leftarrow$  newSolution
11:     else
12:      q  $\leftarrow$  número aleatorio uniforme entre 0 y 1
13:      if q <  $e^{-\text{deltaE}/T}$  then
14:        solution  $\leftarrow$  newSolution
15:      end if
16:     end if
17:     T  $\leftarrow$  T · alpha
18:   end while
19:   return solution
20: end procedure
```

---

Figura 3. Pseudo código del algoritmo de Simulated Annealing

Se diseñaron un total de 4 operadores para explorar el espacio de búsqueda, la función de estos es poder realizar combinaciones sobre una solución inicial para generar sub-soluciones que posteriormente son costeadas hasta que el algoritmo converge. A continuación, se describe cada uno de los operadores y el efecto que producen:

- *Swap-o-matic*: Este operador selecciona dos empleados al azar que tienen habilidades idénticas y están asignados a trabajar el mismo día, pero en turnos distintos. Se intercambian ambos turnos para el mismo día, esto nos permite equilibrar el horario para cumplir con las restricciones gracias a que se ajusta la distribución de la carga entre trabajadores similares.
- *Lucky-guy*: Se elige al azar un empleado, un día y un turno, y cambia el estado de asignación de ese empleado en ese contexto en específico. El cambio es directo, si el empleado debía trabajar, se le asigna en cambio un descanso (eliminar el turno) o viceversa, se le asigna un turno al azar para ese día. Este operador introduce variabilidad en el horario.
- *Time-warp*: Selecciona al azar un empleado junto con dos días diferentes en los cuales se le fue asignado dos turnos diferentes, estos mismo son intercambiados. Este movimiento nos puede ayudar a

explorar diferentes distribuciones de turnos a lo largo de la semana, buscando mejorar la cobertura y reduciendo la fatiga del empleado.

- *Opti-snack*: Se elige un día y un turno al azar, para asignarlo a todos los empleados, independientemente de la programación anterior que hayan tenido. Este es el operador más disruptivo y por lo tanto puede ser útil para salir de un óptimo local.

La función de costo en este caso es un esquema **mono-objetivo**, el costo de asignación de un empleado se suma con los costos generados por las restricciones, las cuales fueron descritas anteriormente en el artículo. Cada que una restricción es violada, esta se suma en un contador para posteriormente multiplicarla por un valor constante de penalización. Matemáticamente se puede expresar como:

$$\begin{aligned}
 f = & \sum_{i=1}^I \sum_{w=1}^W \sum_{d=1}^D k_1 * h_w * x_{iwd} + k_2 \\
 & * \sum_{j=1}^J \sum_{d=1}^D \sum_{t=1}^T \text{máx}(r_{jdt} \\
 & - \sum_{i=1}^I \sum_{w=1}^W e_{wt} * p_{ijwd}, 0) + k_3 \\
 & * \sum_{j=1}^J \sum_{d=1}^D \sum_{t=1}^T \text{máx}(\sum_{i=1}^I \sum_{w=1}^W e_{wt} * p_{ijwd} - r_{jdt}, 0) \\
 & + k_4 * \sum_{i=1}^I \sum_{d=1}^{D-1} \text{máx}(x_{i_{w_1d}} + x_{i_{w_2(d+1)}} - y_{w_1w_2} \\
 & - 1, 0)
 \end{aligned} \tag{17}$$

## PARALELIZACIÓN

Primeramente, se buscó implementar una **paralelización de la función de costo**, cuyo cálculo consistía en varias partes:

1. Se sumaba el total de horas asignadas para cada trabajador, este se multiplicaba por el costo por hora establecido al inicio.
2. Se analizaba si la asignación cumplía con la demanda inicial, si la desviación de esta demanda era positiva se interpretaba como que faltó gente para poder cubrir todos los puestos solicitados, caso contrario cuando la desviación era negativa, se interpretaba como que hay demasiado personal asignado. En ambos casos se aplicaba un costo de penalización.

3. Se verifican las restricciones, cada una de ellas es independiente por lo que se calcula por separado, llevando un contador del total de veces que se quebrantó esta restricción y al final se le aplica una penalización a cada una.

Se buscó realizar el envío de una cantidad  $n$  de trabajadores a cada nodo para que calcularan el primer o segundo paso, siendo  $n$  el total de empleados dividido por el número de nodos a utilizar. Sin embargo, un estudio sobre el tiempo de ejecución de esta sección del código nos mostró que este enfoque no sería efectivo para nuestro caso, ya que el tiempo de procesamiento de estas tareas era muy bajo por lo que la ganancia de paralelizarlo no sería representativa o podría llegar a ser contraproducente [Tabla 6 y 7]. Para futuros pasos se analizará si es posible diseñar un cálculo de la función de costo que reutilice el costo anterior únicamente cambiando la aportación del cambio generado en la nueva solución, para de esta manera no tener que recalcular con cada iteración, y de esta manera mejorar los tiempos de ejecución y reducir la carga de procesamiento.

Actividad	Tiempo (s)
Importación de la data	0.0168
Evaluación de la solución inicial	0.000
Leer la demanda del problema	0.000
Movimiento de búsqueda local	0.15
Evaluación del costo del empleado asignado	0.0156
Evaluación de las restricciones y la demanda	0.0002
Algoritmo converge	173.776

*Tabla 6: Tiempos de ejecución para el algoritmo de Hill Climbing*

Actividad	Tiempo (s)
Importación de la data	0.0168
Evaluación de la solución inicial	0.000
Leer la demanda del problema	0.000
Movimiento de búsqueda local	2.26e-6
Evaluación del costo del empleado asignado	0.0156
Evaluación de las restricciones y la demanda	0.0002
Algoritmo converge	70.816

*Tabla 7: Tiempos de ejecución para el algoritmo de Simulated Annealing*

También se consideró la posibilidad de **dividir el espacio de búsqueda** entre el total de nodos, permitiendo que cada nodo explorara una subregión específica del espacio de búsqueda y combinando los resultados al final para obtener una única solución de asignación. Sin embargo, esta estrategia no fue

factible debido a la naturaleza de las restricciones. Estas restricciones requieren la evaluación de asignaciones previas y futuras para asegurar su legalidad. Además, los operadores de optimización, como el operador de *optisnack*, necesitan tener una visión global del problema y la capacidad de afectar a todos los trabajadores simultáneamente. Esto implica que una división del espacio de búsqueda podría comprometer la integridad y efectividad de los operadores, impidiendo una optimización adecuada.

Se optó por una **paralelización MIR de tipo servidor-cliente**, en la cual el nodo maestro se encargaba únicamente de recibir las respuestas finales, mientras que cada nodo de trabajador realizaba el proceso de optimización de manera independiente. Al finalizar, se obtenía un total de  $n$  respuestas, siendo  $n$  el número total de nodos utilizados. Cada nodo escribía su asignación en documentos *txt* separados, y enviaba al nodo maestro la solución final junto con el costo asociado y un *string* detallando el total de ilegalidades cometidas, empaquetados usando la biblioteca *MPI*.

Cada nodo trabajador enviaba un vector de caracteres que contenía el conteo de cada restricción no cumplida en la asignación, junto con el costo total de la misma. Para empaquetar esta información se utilizó la función *MPI Pack*, almacenando los elementos en el siguiente orden: costo, tamaño del vector y el propio vector. En el nodo maestro, se implementó un bucle que iteraba sobre cada nodo trabajador para recibir la respuesta empaquetada.

Para asegurar que se pudiera recibir el mensaje completo sin conocer previamente su tamaño real, se utilizó la función *MPI Probe* para inspeccionar el mensaje, junto con la función *MPI Get count*, que obtenía el tamaño del mensaje y lo almacenaba en la variable *message size*. Esto permitía generar un buffer con el tamaño adecuado para recibir el mensaje que luego se desempaquetaba. El primer elemento del mensaje correspondía al costo final de la optimización, el segundo valor era el tamaño del vector, y el tercero era el vector de caracteres. Esta información se desempaquetaba utilizando la función *MPI Unpack*.

## INTERFAZ GRÁFICA

La capacidad de interactuar visualmente y modificar las soluciones generadas por los algoritmos de optimización es esencial para su aplicación práctica en entornos operativos. La interfaz gráfica de usuario desempeña un papel fundamental al permitir a los usuarios interactuar con los datos de manera efectiva, brindando una herramienta poderosa para ajustar y mejorar las soluciones propuestas por el optimizador. Actualmente se está mejorando a

la visualización de las restricciones para que sea más intuitivo para el usuario, pero en la Gantt se cumple con la visión de en qué turno o descanso específico se rompe una de las restricciones previamente explicadas. Esto es muy útil para poder visualizar la calidad de la solución generada por el optimizador. La interfaz es realizada con el framework de React, el cual utiliza javascript. Por el momento no tiene funcionalidades para modificar la asignación ni recalcular costos o restricciones, su propósito es meramente visualizar la asignación creada [Figura 4]. Para poder generarla se carga tres archivos csv: la asignación generada, la descripción de horas de cada turno y las restricciones entre turnos que se debe respetar (las demás restricciones también son consideradas, pero estas no se deben cargar como un archivo).

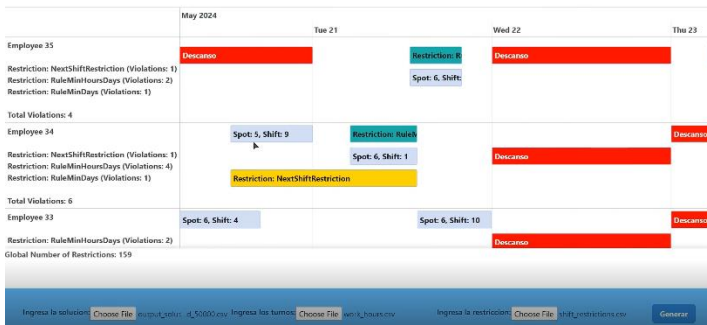


Figura 4: Visualización tipo Gantt de la asignación.

Una asignación se ve plasmada como un recuadro gris dentro del cual viene descrito el spot a donde el empleado fue asignado, así como el turno que debe trabajar. De rojo se agregaron los descansos que tiene. Y las restricciones se visualizan como recuadros encima de la asignación con el detalle de que restricción se rompió. Cuando uno de estos elementos es señalado su color cambia a amarillo.

Además, se agregó un contador de restricciones quebradas a nivel global, así como un contador individual de cada empleado junto con las ilegales que comete su asignación.

## RESULTADO

Para el caso de la paralelización por instancias múltiples independientes podemos analizar la eficacia de cada nodo individual mediante los tiempos de ejecución, en este caso se realizó el estudio con una configuración de 1, 2, 3,

4 y 5 nodos trabajadores que realizan la optimización y analizamos el promedio de los tiempos de entre los nodos. Cada configuración fue realizado un total de 5 veces, para generar promedios.

Considerando que  $T_s$  es el tiempo de ejecución de un solo nodo y que  $T_p$  será el tiempo de ejecución promedio en múltiples nodos, podemos calcular el Speedup con la siguiente formula:

$$S(n) = \frac{n * T_s}{T_p} = n \quad (18)$$

Como cada nodo se ejecuta en una instancia independiente, se debe cumplir que el speedup es proporcional al número de nodos. Para el cálculo de la eficacia se puede realizar de la siguiente manera, en caso de que se cumpla la condición se puede interpretar que cada nodo está siendo utilizado de manera óptima.

$$E(n) = \frac{S(n)}{n} = \frac{n * T_s}{n * T_p} = \frac{T_s}{T_p} \quad (19)$$

Los resultados del HillClimbing muestra que efectivamente en cada configuración de número de nodos se logró una utilización óptima de los nodos [Tabla 8]. Para este algoritmo se utilizó un total de 10,000 iteraciones. Para el caso del Simmulated Annealing se realizó el mismo estudio [Tabla 9], los parámetros utilizados fue una temperatura inicial de 10,000,000 con una tasa de enfriamiento de 0.005 y una temperatura final de 0.1, generando un total de 3675 iteraciones.

Nodos	$T_p$	Costo	Speedup	Eficacia
Serial	212.137	205728.6		
1	160.160	194429.6		
2	154.487	187828.8	1.930	0.964
3	162.863	188629.4	3.051	1.017
4	164.019	189829	4.096	1.024
5	160.167	188829.2	5.000	1.000

Tabla 8: Tabla de resultados de Speedup y Eficacia de las configuraciones de los nodos [1,2,3,4 y 5] para el Hill Climbing.

Nodos	$T_p$	Costo	Speedup	Eficacia
Serial	82.645	726572.25		
1	66.955	638527.2		
2	67.044	616307.6	2.003	1.001

3	69.385	582375.8	3.109	1.036
4	68.105	605396.8	4.069	1.017
5	67.271	599485.8	5.023	1.005

*Tabla 9: Tabla de resultados de Speedup y Eficacia de las configuraciones de los nodos [1,2,3,4 y 5] para el Simulated Annealing.*

De igual manera podemos analizar la ganancia en cuanto a costo promedio que se obtuvo gracias a la ejecución de diversos nodos independientes, a comparación de solo uno, para ello consideramos  $C_s$  como el costo de un solo nodo, y  $C_p$  como el costo promedio para cada configuración de los nodos trabajadores, por lo que la formula sería:

$$R = 1 - \frac{C_p}{C_s} \tag{20}$$

Para el Hill Climbing mantuvimos el mismo número de iteraciones [Tabla 10]. Para el Simulated Annealing utilizando los mismos parámetros [Tabla 11]

Nodos	Costo	Reducción
Serial	205728.6	
1	194429.6	
2	187828.8	0.034
3	188629.4	0.030
4	189829	0.024
5	188829.2	0.029

*Tabla 10: Tabla de resultados del Costo promedio obtenido de las configuraciones de los nodos [1,2,3,4 y 5] para el Hill Climbing*

Nodos	Costo	Reducción
Serial	726572.25	
1	638527.2	
2	616307.6	0.035
3	582375.8	0.088
4	605396.8	0.052
5	599485.8	0.061

*Tabla 11: Tabla de resultados del Costo promedio obtenido de las configuraciones de los nodos [1,2,3,4 y 5] para el Simulated Annealing*



## CONCLUSIONES

En este caso los tiempos de ejecución son muy consistentes entre diferentes nodos, todos obteniendo una eficiencia cercana al 1, lo que indica que la paralelización por MIR se realizó de manera correcta y los recursos están siendo utilizados correctamente. Esto tiene implicaciones en la escalabilidad del número de nodos para poder manejar más instancias sin que haya riesgo de una disminución en la eficiencia, considerando siempre que el trabajo se distribuya equitativamente y no haya interferencias significativas entre las tareas ejecutadas en cada nodo.

En cuanto a los promedios de los costos obtenidos, se puede apreciar que el uso de Simulated Annealing paralelizado nos permitió explorar más el espacio de búsqueda y evitar caer en mínimos locales, como se puede apreciar con el Hill Climbing donde la reducción del costo obtenido en serial contra las configuraciones de la paralelización apenas hay diferencias.

## FUTUROS TRABAJOS

Para futuros trabajos se buscará analizar el efecto de iniciar el algoritmo de optimización tomando una solución inicial generada de manera aleatoria, de esta manera evitamos el uso de cargar una solución prefabricada que nos podría llevar a un mínimo local. También se plantea el utilizar algoritmos miméticos para generar una población de soluciones más grande y poder expandir nuestro espacio de búsqueda, se opta estos algoritmos como un siguiente paso factible ya que se ha demostrado su eficacia para resolver problemas de programación. En este caso también se optaría por una metodología de paralelización MIR de metaheurísticas de trayectoria como las exploradas en este estudio, como primera opción, para realizar una búsqueda local de las soluciones de descendencia. Para esto se buscará hacer uso de una cantidad grande de cómputo para poder manejar esta búsqueda exhaustiva y generación grande poblacional.

En cuanto a la interfaz gráfica, se buscará mejorar la visualización de las restricciones para que sea más adecuada al usuario, junto con la implementación de funcionalidades que nos permitan interactuar con la asignación, y de esa manera, poder manualmente solucionar los errores que presenta, entre otros detalles que puede ser mejorados.

## REFERENCIAS

- [1] Ormecí, E. L., Salman, F. S., and Yücel, E. (2014). Staff rostering in call centers providing employee transportation. *Omega*, 43:41–53.

- [2] Henderson, S., Mason, A., Ziedins, I., and Thomson, R. (1999). A heuristic for determining efficient staffing requirements for call centres.
- [3] Chuin Lau, H. (1996). On the complexity of manpower shift scheduling. *Computers Operations Research*, 23(1):93–102.
- [4] Ernst, A., Jiang, H., Krishnamoorthy, M., and Sier, D. (2004). Staff scheduling and rostering: A review of applications, methods and models. *European Journal of Operational Research* 153:3–27.
- [5] Alba, E. (2005). *Parallel Metaheuristics. A New Class Of Algorithms*. Wiley-Interscience.
- [6] Kletzander, L. and Musliu, N. (2019). Solving the general employee scheduling problem. *Computers Operations Research*, 113:104794.
- [7] Ramli, R., Ahmad, S., Abdul Rahman, S., and Wibowo, A. (2020). A tabu search approach with embedded nurse preferences for solving nurse rostering problem. *International Journal for Simulation and Multidisciplinary Design Optimization*, 11:10.
- [8] Azencott, R. (1992). *Simulated Annealing: Parallelization Techniques*. John Wiley and Sons.
- [9] Greening, D.R. (1995). *Simulated annealing with errors*. PhD Dissertation, University of California, Los Angeles.
- [10] Leite, J.P.B., and Topping, B.H.V., (1999). "Parallel simulated annealing for structural optimization. *Computers and Structures*, 73, 545-564.
- [11] Meise, C. (1998). "On the convergence of parallel simulated annealing". *Stochastic Processes and their Applications*, 76, 99-115.
- [12] Bongiovanni, G., Crescenzi, P., and Guerra, C., (1995). "Parallel simulated annealing for shape detection". *Computer Vision and Image Understanding*, 61(1), 60- 19.
- [13] Vales-Alonso, J., Fernandez J., Gonzalez-Castano, F.J., and Cabarello, A., (2003). "A parallel optimization approach for controlling allele diversity in conversation schemes". *Mathematical Biosciences*, 183 (2), 161-173.
- [14] Bhandarkar, S.M., Machaka, S., Chirravuri, S., and Arlond, J., (1998). "Parallel computing for chromosome reconstruction via ordering of DNA sequences". *Parallel Computing*, 24(12-13), 1177-1204.
- [15] Chu, K.W., Deng, Y., and Reinitz, J., (1999). "Parallel simulated annealing by mixing of states". *Journal of Computational Physics*, 148 (2), 646-662.
- [16] Bevilacqua, A., (2002). "A methodological approach to parallel simulated annealing on an SMP system". *Journal of Parallel and Distributed Computing*, 62(10), 1548-1570.
- [17] Brown, C., (1994). *Unix Distributed Computing*. Prentice Hall, London.

- [18] Shiau, J.-Y., Huang, M.-K., and Huang, C.-Y. (2020). A hybrid personnel scheduling model for staff rostering problems. *Mathematics*, 8(10).

# MODELADO QSAR DE LA EFICIENCIA EN LA INHIBICIÓN DE LA CORROSIÓN MEDIANTE FÁRMACOS Y FITOQUÍMICOS

**Alan Miralrio, Carlos Beltrán Araceli Espinoza Vázquez  
Pérez, Andrés Antonio  
Abundis Serrano, Mario  
Villada**

*Tecnológico de Monterrey, Escuela de  
Ingeniería y Ciencias, Ave. Eugenio Garza  
Sada 2501, Monterrey 64849, Nuevo León,  
México.*

*Email: [miralrio@tec.mx](mailto:miralrio@tec.mx)*

*web page: <https://tec.mx/es>*

*Unidad Anticorrosión, Instituto de  
Ingeniería, Universidad Veracruzana,  
Boca del Río 94292, Veracruz, México  
Email: [araespinoza@uv.mx](mailto:araespinoza@uv.mx),  
web page: <https://www.uv.mx/>*

**Resumen.** En este trabajo, se compararon varias metodologías con el fin de obtener modelos claros, portables y precisos para la predicción de la eficiencia de inhibición de la corrosión. Basándose en trabajos previos, se implementaron modelos como un método robusto para obtener modelos no lineales portables y para determinar las variables más importantes para analizar el fenómeno de la inhibición de la corrosión. También, se asumió una aproximación basada en la teoría de ácido-base duro-blando de Pearson (HSAB, por sus siglas en inglés). Mediante el paradigma de la relación cuantitativa estructura-actividad (QSAR, por sus siglas en inglés), comúnmente empleado en el desarrollo de nuevos medicamentos, se estudiaron 650 fármacos diferentes. Se calcularon diversos descriptores químico-cuánticos mediante el método de amarre fuerte basado en funcionales de la densidad (DFTB, por sus siglas en inglés). Los descriptores consideraron principalmente las propiedades electrónicas y energéticas de las moléculas de interés, como son: energía de ionización (I), afinidad electrónica (A), electronegatividad ( $\chi$ ), dureza ( $\eta$ ), electrofilicidad ( $\omega$ ) y fracción de electrones transferidos ( $\Delta N$ ). Además, se utilizó el método de mínimos cuadrados ordinarios (OLS, por sus siglas en inglés) para obtener modelos lineales con un número reducido de variables y proponer reglas generales basadas en información sin procesar. Finalmente, se empleó la inteligencia artificial IBM Watson, basada en la regresión de árboles de decisión, como una alternativa robusta pero no portátil y altamente parametrizada a los enfoques matemáticos convencionales. La comparación entre los modelos se realizó en términos de métricas como el error porcentual absoluto medio (MAPE) y la raíz del error cuadrático medio (RMSE). En general, los modelos

con un número reducido de variables y términos de hasta segundo orden han mostrado un rendimiento mejorado. Además, se determinó el papel determinante de la energía del orbital molecular más alto ocupado (HOMO, por sus siglas en inglés).

**Palabras clave:** QSAR, corrosión, OLS, inhibidores de corrosión, IBM Watson, fármacos.

## INTRODUCCIÓN

Entre los metales, el acero es el más empleado en diversas industrias. Algunas de estas son: petrolífera, alimentaria, energética, química y de la construcción. Así, el hierro y sus aleaciones son de alto interés por sus excelentes propiedades mecánicas, al ser dúctil, duradero y resistente<sup>1</sup>. Por todo lo anterior, proveer soluciones innovadoras que ayuden a reducir los efectos nocivos de la corrosión en metales es un tema prioritario para el mundo y en particular para México. Sin embargo, los tratamientos anticorrosivos más comunes pueden estar relacionados a problemáticas medioambientales, debido al uso de sustancias tóxicas<sup>1</sup>.

Una estrategia recurrentemente utilizada es emplear inhibidores de corrosión en su superficie. Un inhibidor de corrosión es una sustancia que, agregada en pequeñas cantidades a la superficie del metal, reduce la acción del medio corrosivo sobre el metal. En breve, las moléculas del inhibidor forman una película protectora<sup>1</sup>. La efectividad de protección de estos inhibidores depende en gran medida del tipo de superficie, del tipo de inhibidor y de condiciones propias del ambiente en el cual se aplica el tratamiento. En este contexto, una variedad de compuestos orgánicos ha sido considerados como efectivos en inhibir la corrosión. Mayormente, los compuestos descritos como inhibidores orgánicos altamente eficientes contienen átomos de nitrógeno, oxígeno y azufre. Además, las moléculas inhibitoras son ricas en electrones  $\pi$  asociados a enlaces triples, enlaces conjugados dobles o anillos aromáticos<sup>1</sup>.

En años recientes, se han estudiado inhibidores de extractos de plantas y medicamentos de uso común<sup>2-5</sup>. Al existir una inmensa cantidad de fitoquímicos (sustancias obtenidas de las plantas) y fármacos disponibles para ser probados como inhibidores de la corrosión en metales, se deben definir estrategias claras para estudiarlos<sup>5</sup>. De esta forma, el presente proyecto pretende emplear herramientas usadas en el diseño y predicción de nuevos fármacos, como el paradigma relación cuantitativa estructura-propiedad (QSAR, por sus siglas en inglés) pero empleándolas para proponer nuevos

inhibidores de la corrosión. Así, es posible darles un valor agregado a los excedentes producidos por la industria farmacéutica nacional y/o darle un destino final a las sustancias que hayan alcanzado su caducidad.

## **OBJETIVOS**

### **Objetivos generales**

Determinar la importancia de los descriptores comúnmente empleados en la obtención de nuevos fármacos, mediante QSAR, en su efecto como inhibidores de la corrosión en metales.

### **Objetivos particulares**

- Calcular los descriptores químico-cuánticos descritos en la teoría HSAB de Pearson, mediante aproximaciones en DFTB, para al menos 1,500 fármacos y fitoquímicos.
- Consolidar la base de datos con valores de energía de enlace superficie-ligando obtenidas mediante dinámica molecular.
- Obtener un modelo, mediante técnicas de ajuste de datos como el análisis de regresión multivariado, que tenga poder de predicción de la eficiencia en la inhibición de la corrosión y de la energía de enlace superficie-ligando.

## **METODOLOGÍA**

### **Descriptores químico-cuánticos**

Los estudios teóricos pueden elucidar la naturaleza del efecto inhibidor de las moléculas orgánicas. Además, las aproximaciones teóricas han demostrado ser altamente efectivas para elucidar las interacciones metal-inhibidor, a una escala atómica que comúnmente no es accesible mediante técnicas experimentales, en particular en el caso del hierro<sup>6</sup>. Las metodologías basadas en la teoría del funcional de la densidad (DFT, por sus siglas en inglés) es la más comúnmente empleada, ya que obtiene resultados precisos con un costo computacional adecuado<sup>7</sup>. En una descripción químico-cuántica, las propiedades químicas de una molécula están fuertemente determinadas por sus orbitales moleculares de frontera, así como sus energías, estos son: orbital molecular más alto ocupado (HOMO, por sus siglas en inglés) y orbital molecular más bajo desocupado (LUMO, por sus siglas en inglés). Los

orbitales de frontera proveen de información acerca de la reactividad química, así como propiedades energéticas de la molécula en cuestión<sup>7</sup>.

Otros parámetros que se han propuesto para analizar la interacción entre los inhibidores de la corrosión orgánicos y las superficies metálicas son aquellos propuestos dentro de la teoría ácido-base duro-blando (HSAB, por sus siglas en inglés) de Pearson<sup>8</sup>. Estos se basan en el potencial de ionización vertical (I) y afinidad electrónica vertical (A). Según el teorema de Koopmans, ambas se pueden obtener de las energías del HOMO y LUMO, tal que:  $I = -E_{\text{HOMO}}$  y  $A = -E_{\text{LUMO}}$ . La combinación de ambas da lugar a las relaciones que se emplean en la teoría HSAB, la cual predice que dos sustancias interactuarán más fácilmente si una es un ácido “duro” y la otra una base “dura” o si uno es un ácido “blando” y la otra una base “blanda”<sup>8</sup>. La forma de determinar la dureza absoluta ( $\eta$ ) de un compuesto es mediante  $\eta = (I-A)/2$ .

Otras propiedades pueden obtenerse dentro de la teoría HSAB, como electronegatividad absoluta  $\chi = (I+A)/2$  y la electrofilicidad  $\omega = (I+A)^2(8*(I+A))^{-1}$ <sup>9</sup>. El estudio no se limitará a estas definiciones. Una última relación, en el marco de HSAB, da información valiosa sobre el actuar de la molécula orgánica con la superficie metálica. Esta es la fracción de electrones transferidos  $\Delta N = (\chi_{\text{Metal}} - \chi_{\text{Inhibidor}}) / [2*(\eta_{\text{Metal}} + \eta_{\text{Inhibidor}})]$ ; donde se emplean las electronegatividades y durezas absolutas de la superficie metálicas y el inhibidor. En este apartado, los estudios de Lukovits y colaboradores, dan como resultado que a mayor  $\Delta N$  es mayor la eficiencia en la inhibición<sup>10,11</sup>.

### **Amarre fuerte basado en funcional de la densidad**

Aunque actualmente DFT es la metodología más empleada para el estudio de las moléculas orgánicas como inhibidores de la corrosión, los cálculos para moléculas complejas pueden requerir de tiempos de cómputo largos, incluso en equipos de HPC<sup>7</sup>. La aproximación de amarre fuerte basada en teoría del funcional (DFTB, por sus siglas en inglés) es una propuesta que ayuda a disminuir el tiempo de cómputo en cálculos intensivos, escalando en un orden  $N^2$ , lo que permite calcular sistemas complejos en equipos de cómputo menos robustos y en un tiempo computacional razonable<sup>12,13</sup>.

El marco teórico de DFTB se basa, al igual que en DFT, en el uso de la densidad electrónica  $\rho(\mathbf{r})$  para obtener toda la información del sistema. En una primera aproximación, la densidad electrónica  $\rho(\mathbf{r}) = \rho_0(\mathbf{r}) + \delta\rho(\mathbf{r})$  de una molécula se puede proponer como la superposición lineal de densidades electrónicas atómicas  $\rho_0(\mathbf{r})$  más una pequeña perturbación  $\delta\rho(\mathbf{r})$ . Así, la energía total del sistema  $E[\rho] = E[\rho_0 + \delta\rho]$  se puede expandir en una serie de Taylor hasta un tercer orden, denominado DFTB3<sup>12,13</sup>.

$$E^{DFTB3}[\rho_0 + \delta\rho] = E^0[\rho_0] + E^1[\rho_0, \delta\rho] + E^2[\rho_0, (\delta\rho)^2] + E^3[\rho_0, (\delta\rho)^3] \quad (1)$$

Donde los términos son :

$$E^0[\rho_0] = 1/2 \sum_{AB} (Z_A Z_B / R_{AB}) - 1/2 \iint (\rho_0(\mathbf{r}) \rho_0(\mathbf{r}') / |\mathbf{r} - \mathbf{r}'|) d\mathbf{r} d\mathbf{r}' - \int V^{XC}[\rho_0] \rho_0(\mathbf{r}) d\mathbf{r} + E^{XC}[\rho_0], \quad (2)$$

$$E^1[\rho_0, \delta\rho] = \sum_i n_i \langle \psi_i | H[\rho_0] | \psi_i \rangle, \quad (3)$$

$$E^2[\rho_0, (\delta\rho)^2] = 1/2 \iint [1/|\mathbf{r} - \mathbf{r}'| + \delta^2 E^{XC}[\rho] / (\delta\rho(\mathbf{r}) \delta\rho(\mathbf{r}') |_{\rho_0})] \delta\rho(\mathbf{r}) \delta\rho(\mathbf{r}') d\mathbf{r} d\mathbf{r}',$$

$$E^3[\rho_0, (\delta\rho)^3] = 1/6 \iiint (\delta^3 E^{XC}[\rho] / (\delta\rho(\mathbf{r}) \delta\rho(\mathbf{r}') \delta\rho(\mathbf{r}'') |_{\rho_0}) \delta\rho(\mathbf{r}) \delta\rho(\mathbf{r}') \delta\rho(\mathbf{r}'') d\mathbf{r} d\mathbf{r}' d\mathbf{r}'' \quad (4)$$

Los términos de DFTB3 se pueden obtener de diversas fuentes, el término de orden  $E^0[\rho_0]$  se obtiene de la densidad electrónica de referencia. Los términos de orden superior,  $E^2[\rho_0, (\delta\rho)^2]$  y  $E^3[\rho_0, (\delta\rho)^3]$ , se obtienen mediante cálculos auto consistentes (SCC, por sus siglas en inglés). Obtener estos términos, como en el caso de DFT tradicional, requiere un poder de cómputo elevado debido al alto número de cálculos a realizar<sup>12,13</sup>.

Otras limitaciones que imponen los métodos DFTB es que requieren parametrizaciones para los pares de átomos interactuantes. Sin embargo, actualmente no existen parametrizaciones para toda la tabla periódica, por lo que se restringe su uso como alternativa al DFT tradicional. Estos parámetros consisten en dos partes, parámetros de la parte electrónica del modelo DFTB, así como otros de los términos de repulsión entre dos centros. El conjunto de estos parámetros, calculados previamente para cada par de átomos presente en la molécula, reduce el tiempo de cómputo respecto de DFT. Como ventajas, DFTB permite obtener valores, comparables a DFT tradicional, de los niveles energéticos  $E_{HOMO}$  y  $E_{LUMO}$ . En consecuencia, los descriptores propuestos por Pearson en HSAB, como dureza  $\eta$ , electrofilicidad  $\omega$ , fracción de electrones transferidos  $\Delta N$ , entre otros. Además, diferentes aproximaciones de DFTB se encuentran implementados en el código del software paralelizado llamado DFTB+<sup>12,13</sup>. Esta es una herramienta de simulación en química cuántica de software libre, la cual es capaz de realizar dinámicas moleculares para simular el sistema en diferentes condiciones de temperatura, volumen y presión. Así, es posible modelar las superficies del metal en interacción con la molécula orgánica.



## **Paradigma estructura-propiedad**

Finalmente, como se explicó anteriormente, las propiedades electrónicas y energéticas de las moléculas orgánicas pueden relacionarse con su eficiencia como inhibidores de la corrosión, llamado enfoque QSAR. Para obtener estos modelos, QSAR retoma múltiples herramientas del análisis estadístico de datos y ciencia de datos. Actualmente, los métodos de QSAR se basan en técnicas para ajustar modelos, lineales o no lineales, a un gran volumen de descriptores calculados para las moléculas aisladas. Entre estos destacan los análisis de regresión múltiple (MRA) y más recientemente los modelos autoregresivos con entradas exógenas (ARX, por sus siglas en inglés). Este último extensible a modelos no lineales<sup>14</sup>. Recientemente, el grupo de investigación ha reportado modelos QSAR para la predicción de la inhibición de la corrosión empleando ARX y Watson, siendo esta última la inteligencia artificial de IBM<sup>15</sup>.

## **RELACIÓN CON PROBLEMAS NACIONALES**

México tiene una fuerte industria farmacéutica, la cual en 2019 produjo 21,476 millones de dólares en fármacos<sup>16</sup>. Sin embargo, posee un rezago en cuanto a la manipulación de los fármacos caducos y excedentes, debido a las malas prácticas en su disposición final y la nula posibilidad de reúso<sup>17</sup>. Por lo anterior, se puede considerar que estas sustancias representan un potencial peligro medioambiental para el país. Lo anterior también puede aplicarse en el caso de desechos agrícolas, los cuales podrían emplearse por sus fitoquímicos.

El presente trabajo pretende explorar una aplicación que, en principio, permitiría extender la vida útil de un fármaco al emplearlo como inhibidor de la corrosión en metales. De esa forma, los excedentes producidos podrían reincorporarse en una nueva vertiente, al igual que se estimularía la recuperación de fármacos que, de otra forma, afectarían el medio ambiente si no se dispone de ellos adecuadamente.

## **PROPUESTA DE SOLUCIÓN**

El presente proyecto pretende implementar las herramientas usadas en el diseño y predicción de nuevos fármacos, dentro del paradigma QSAR, pero empleándolas para proponer nuevos inhibidores de la corrosión altamente

eficientes. De esta forma, es posible darles un valor agregado y vida extendida a los excedentes producidos por la industria farmacéutica nacional. Para este fin, se propuso emplear un esquema QSAR que empleara los descriptores propuestos dentro de la teoría ácido-base duro-blando (HSAB, por sus siglas en inglés) de Pearson<sup>8</sup>.

Los descriptores en HSAB se basan en el potencial de ionización vertical (I) y afinidad electrónica vertical (A). La combinación de ambas da lugar al principio HSAB, el cual dice que dos sustancias interactuarán más fácilmente si una es un ácido “duro” y la otra una base “dura” o si uno es un ácido “blando” y la otra una base “blanda”<sup>8</sup>. La forma de cuantificar la dureza absoluta ( $\eta$ ) de un compuesto es mediante  $\eta = (I-A)/2$ . Otras propiedades comunes que pueden obtenerse dentro de la teoría HSAB, son: electronegatividad absoluta  $\chi = (I+A)/2$  y la electrofilicidad  $\omega = (I+A)^2(8*(I+A))^{-1}$ <sup>9</sup>.

La teoría del funcional de la densidad o la aproximación de amarre fuerte basada en teoría del funcional (DFTB, por sus siglas en inglés) son dos metodologías que pueden emplearse para obtener los descriptores de HSAB. Sin embargo, DFTB fue propuesta en este proyecto ya que ayuda a disminuir el tiempo de cómputo en cálculos intensivos, lo que permite calcular sistemas complejos en equipos de cómputo menos robustos y en un tiempo computacional razonable<sup>12,13,18</sup>.

## RESULTADOS Y LOGROS

Este estudio se presentó una exploración exhaustiva de tres distintas familias de modelos de regresión con el objetivo de predecir el potencial de inhibición de la corrosión de medicamentos comerciales como inhibidores orgánicos de la corrosión mediante asociaciones de relación cuantitativa estructura-actividad (QSAR). El análisis abarcó un conjunto de datos de 60 medicamentos comerciales, combinando descriptores fisicoquímicos derivados experimentalmente con propiedades calculadas energéticamente utilizando metodologías químicas cuánticas. Se empleó una estrategia de validación cruzada de “5-fold” para construir todos los modelos, garantizando robustez y generalización. Para facilitar una comparación equitativa entre los modelos propuestos, se introdujo un conjunto de validación externo que comprendía 10 medicamentos comerciales adicionales.

Se conformó una base de datos, en formato CSV, que contiene información básica de 650 fármacos de uso común en México. Los datos iniciales fueron: Nombre común, nombre según la IUPAC, número CAS, fórmula química, identificador canónico SMILE, identificador INCHL y peso molecular. En particular, el canónico SMILE sirvió para proponer la información inicial de la estructura tridimensional de las moléculas bajo estudio.

Los siguientes datos que también se incluyeron en la base de datos se usan comúnmente como descriptores en análisis QSAR. Como se explicó anteriormente, las propiedades fisicoquímicas, electrónicas y energéticas de las moléculas orgánicas pueden relacionarse con alguna actividad biológica. Así, los descriptores que se incluyeron fueron aquellos que se emplean comúnmente en el desarrollo de fármacos, estos son: coeficiente de solubilidad LogS, pKa, coeficiente de partición LogP, área de la superficie polar, polarizabilidad y sitios aceptores o donadores de hidrógeno.

En tercera instancia, se incluyeron los datos obtenidos de los cálculos químico-cuánticos basados en DFTB. Los cálculos partieron de las estructuras iniciales extraídas del canónico SMILE. Seguidamente, se optimizaron las estructuras con el código DFTB+ en los servidores del Laboratorio de Supercómputo del Bajío. De los fármacos en sus estructuras en los estados base se extrajo el valor de la energía de su HOMO y también del LUMO. Mediante el teorema de Koopmans, se calcularon otros descriptores, los cuales también se incluyeron en la base de datos. Estos son: energía de ionización, afinidad electrónica, electronegatividad, dureza, electrofiliidad y finalmente fracción de electrones transferidos al hierro.

Se empleó la base de datos para investigar diversos modelos matemáticos ARX para la predicción de la inhibición de la corrosión en aleaciones de hierro. Las principales rutinas, necesarias de ajuste del modelo ARX, se realizaron mediante la paquetería Matlab. Primeramente, se desarrollaron experimentos que empleaban 80% de los registros para el conjunto de ajuste y el 20% para el conjunto de prueba. Sin embargo, los resultados más favorables se obtuvieron con un conjunto 80-20. Además de lo anterior, se pudo corroborar que no todos los descriptores contribuyen en la misma medida a la predicción de la inhibición de la corrosión. Gracias a esto, se pudo encontrar que el pKa y todas las propiedades químico-cuánticas obtenidas con DFTB son suficientes para formular modelos matemáticos lineales que obtuvieron predicciones favorables.

La investigación reveló que cada enfoque exhibía fortalezas y limitaciones

inherentes. Notablemente, el modelo NARX-FROLS facilitó la reducción de dimensionalidad de diez descriptores iniciales a solo cinco variables fundamentales:  $E_{\text{HOMO}}$ ,  $E_{\text{LUMO}}$ ,  $pK_a$ ,  $\omega$  y  $\Delta N$ . Se destaca el papel crítico de los atributos energéticos y electrónicos en la racionalización del comportamiento de inhibición de la corrosión de los compuestos orgánicos. También, se subrayó la influencia del  $pK_a$ , que indica la predisposición de una sustancia orgánica a disociarse en medio acuoso. Importante es que el modelo NARX de segundo orden ofreció perspectivas sobre la energía óptima de  $E_{\text{HOMO}}$  ( $\sim -5.6202$  eV a nivel de teoría DFTB3-LJ/3OB), mejorando la comprensión del comportamiento molecular. Además, el uso de la Regresión de Árboles Extra a través de Watson AutoAI de IBM surgió como un modelo altamente preciso y exacto en este estudio. Las métricas de validación externa, incluyendo MAPE, MSE y RMSE, demostraron resultados prometedores de aproximadamente 5.93, 44.01 y 6.623, respectivamente. La culminación de este trabajo implicó estimar la eficiencia de inhibición de la corrosión utilizando los modelos QSAR óptimos para un conjunto de 750 medicamentos. Esta evaluación confirmó el consenso entre las 10 mejores sustancias, enfatizando el potencial para futuras investigaciones exhaustivas sobre las propiedades de inhibición de la corrosión de estos candidatos.

## REFERENCIAS

- [1] V. S. Sastri, *Green Corrosion Inhibitors: Theory and Practice*, John Wiley and Sons, New Jersey, **2011**.
- [2] A. Espinoza Vázquez, L. A. López Reséndiz, I. A. Figueroa, F. J. Rodríguez Gómez, M. Figueroa, D. Ángeles Beltrán, M. Castro and A. Miralrio, *Journal of Adhesion Science and Technology*, **2021**, 35, 873–899.
- [3] A. Espinoza-Vázquez, F. J. Rodríguez-Gómez, R. Mata, A. Madariaga-Mazón and D. Ángeles-Beltrán, *J Solid State Electrochem*, **2017**, 21, 1687–1697.
- [4] A. Espinoza-Vázquez, F. J. Rodríguez-Gómez, I. A. Figueroa-Vargas, A. Pérez-Vásquez, R. Mata, A. Miralrio, R. Galván-Martínez, M. Castro and R. Orozco-Cruz, *IJMS*, **2022**, 23, 3130.
- [5] A. Miralrio and A. Espinoza Vázquez, *Processes*, **2020**, 8, 942.
- [6] P. Limon, A. Miralrio and M. Castro, *Computational and Theoretical Chemistry*, **2018**, 1129, 37–47.
- [7] I. N. Levine, *Química cuántica*, Pearson Educación, **2001**.
- [8] R. G. Pearson, *Journal of the American Chemical society*, **1963**, 85, 3533–3539.
- [9] P. K. Chattaraj and S. Giri, *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.*, **2009**, 105, 13.

- [10] A. Kokalj, *Corrosion Science*, **2021**, 180, 109016.
- [11] I. Lukovits, E. Kálmán and F. Zucchi, *CORROSION*, **2001**, 57, 3–8.
- [12] B. Aradi, B. Hourahine and Th. Frauenheim, *J. Phys. Chem. A*, **2007**, 111, 5678–5684.
- [13] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshayé, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irlé, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W. -z. Yu and T. Frauenheim, *J. Chem. Phys.*, **2020**, 152, 124101.
- [14] C. Beltran-Perez, H.-L. Wei and A. Rubio-Solis, *Int. J. Autom. Comput.*, **2020**, 17, 55–70.
- [15] C. Beltran-Perez, A. A. A. Serrano, G. Solís-Rosas, A. Martínez-Jiménez, R. Orozco-Cruz, A. Espinoza-Vázquez and A. Miralrio, *IJMS*, **2022**, 23, 5086.
- [16] G. P. Zazueta, *Industria Farmacéutica: Unidad de Inteligencia de Negocios*, México DF: ProMéxico, **2013**.
- [17] C. P. Hernández-Barrios, G. Fernández-Villagómez and J. Sánchez-Gómez, *Centro Nacional de Prevención de Desastres (CENAPRED), Instituto Nacional de Ecología (INE). México*, **1995**.
- [18] M. Wahiduzzaman, A. F. Oliveira, P. Philipsen, L. Zhechkov, E. van Lenthe, H. A. Witek and T. Heine, *Journal of Chemical Theory and Computation*, **2013**, 9, 4006–4017.

# ALMACENAMIENTO DE HIDRÓGENO EN GRAFENO NITROGENADO DECORADOS CON CÚMULOS METÁLICOS

**Alan Miralrio, Eduardo  
Rangel**

*Tecnológico de Monterrey, Escuela de  
Ingeniería y Ciencias, Ave. Eugenio Garza  
Sada 2501, Monterrey 64849, Nuevo León,  
México.*

*Email: [eduardorc@tec.mx](mailto:eduardorc@tec.mx)*

*web page: <https://tec.mx/es>*

**Resumen.** La teoría funcional de la densidad se utilizó para estudiar la adsorción de hidrógeno en pequeños cúmulos de paladio-platino  $\text{Pd}_{13}$ ,  $\text{Pd}_{12}\text{Pt}$  y  $\text{Pt}_{12}\text{Pd}$  soportados sobre grafeno dopado con nitrógeno. La transferencia de carga y la fuerte unión (hasta cuatro veces mayor que la energía de unión del grupo de Pd soportado sobre grafeno) entre la superficie de grafeno-nitrógeno y el cúmulo evitan el desprendimiento del metal y conducen a tres tipos de estados de adsorción de hidrógeno. El primer tipo es un hidrógeno molecular fisisorbido, el segundo es un estado activado de  $\text{H}_2$  sin barreras de adsorción donde el enlace H-H está relajado y el tercer tipo es un estado disociado.

**Palabras clave:** grafeno-nitrogenado, cúmulos, adsorción, hidrógeno, paladio-platino.

## 1 INTRODUCCIÓN

La adsorción de hidrógeno implica la activación y disociación del hidrógeno molecular, que son indispensables para almacenar hidrógeno como combustible para vehículos y también para aplicaciones de pilas de combustible<sup>1-2</sup>. La naturaleza del soporte (superficie) y del catalizador tiene una influencia importante en el rendimiento catalítico, en términos de adsorción, disociación y desorción de hidrógeno<sup>3</sup>. Así, las nanopartículas de metales de transición representan los catalizadores más prometedores para la activación, disociación y, en consecuencia, para la adsorción del hidrógeno.

En este caso, los catalizadores de Pt y Pd se pueden utilizar para una amplia gama de reacciones como la hidrogenación, la deshidrogenación y la deshalogenación. La quimisorción disociativa de hidrógeno en grupos de Pt y Pd implica fuertes enlaces entre los grupos de Pt y Pd y el hidrógeno<sup>3</sup>. Además de los catalizadores; el soporte es importante porque condiciona el sitio activo en la reactividad<sup>3</sup>. Además son indispensables para la eficiencia y estabilidad de los sistemas catalíticos, incluso podrían funcionar como catalizadores en sí mismos. El soporte de grafeno dopado con nitrógeno ha llamado la atención debido al hecho de que el dopaje con nitrógeno es capaz de modificar las propiedades de los sistemas de soporte de carbono para diversas aplicaciones de interés<sup>4</sup>. Recientemente estudiamos la interacción entre grafeno-nitrogenado y pequeños cúmulos de Pd ( $Pd_n$ ,  $n = 1-4$ ) para mejorar almacenamiento de hidrógeno, a través de un mecanismo de spillover<sup>1,4</sup>. Encontramos barreras energéticas de migración de 0.5 a 0.8 eV (11.53 a 18.45 kcal/mol), lo que sugiere que la migración de hidrogeno adsorbido en los cúmulos de Pd al soporte de N-grafeno puede ocurrir espontáneamente a temperatura ambiente. El principal objetivo de este artículo es estudiar la adsorción de los cúmulos metálicos de paladio-platino  $Pd_{13}$ ,  $Pd_{12}Pt$  y  $Pt_{12}Pd$  en el sustrato de grafeno-nitrógeno.

## 2 OBJETIVOS

### Objetivos generales

Estudiar la adsorción de los cúmulos metálicos de paladio-platino  $Pd_{13}$ ,  $Pd_{12}Pt$  y  $Pt_{12}Pd$  en el sustrato de grafeno-nitrógeno.

## 3 METODOLOGÍA

Se utilizó la teoría funcional de la densidad dentro de la aproximación de gradiente generalizado (GGA) y el código Quantum Espresso. Usamos pseudopotenciales y como el funcional de intercambio y correlación la aproximación de Perdew-Burke-Ernzerhof (PBE)<sup>5</sup>.

## 4 RESULTADOS Y LOGROS

Primero optimizamos las superceldas que modelan al grafeno con defectos. Los defectos que se estudiaron son los piridínicos, pirrólicos y grafiticos. En

la Figura 1 se muestra cada uno de ellos (color amarillo carbonos; color negro nitrógenos)

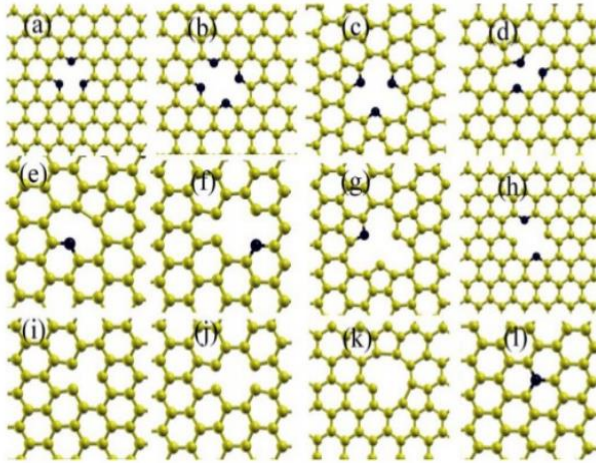


Figura 1 – Superficies de grafeno con diferentes porcentajes de nitrógeno optimizadas. En este contexto,  $N_xV_y$  representa el defecto.  $x$  es el número de átomos de carbono sustituidos por  $N$  átomos de nitrógeno y  $y$  es el número de  $V$  vacancias.

Nuestros resultados muestran que los átomos de nitrógeno pueden estabilizar la formación de vacancias simples, dobles y triples en el grafeno. Generalmente, parece que la energía de formación de  $N_xV_y$  disminuye con el número de átomos de nitrógeno. Estos resultados sugieren que la concentración de los defectos  $N_3V_1$  y  $N_4V_2$  piridínicos, y  $N_3V_3$  y  $N_3V_1$  pirrólicos, predominará más que otros defectos en condiciones experimentales.

Sistema	Energía de formación (eV)	Momento magnético
$N_3V_3$ pirrólico	8.07	
$N_3V_1$ pirrólico	5.52	0.92
$N_4V_2$ piridínico	3.65	0
$N_3V_1$ piridínico	3.3	0.35



*Tabla 1 : Energías de formación y momentos magnéticos de los sistemas de grafeno nitrogenado mas estables.*

El sistema estudiado en este trabajo implica inicialmente un sustrato piridínico  $N_4V_2$  (contenido de nitrógeno del 8,3%) y un conjunto de  $Pd_{13}$ ,  $PdPt_{12}$  y  $PtPd_{12}$  como catalizadores. En particular, el contenido de nitrógeno en nuestro modelo es comparable a los resultados experimentales. Encontramos que las estructuras son estables durante la optimización. La Figura 2 muestra la optimización de los sistemas.

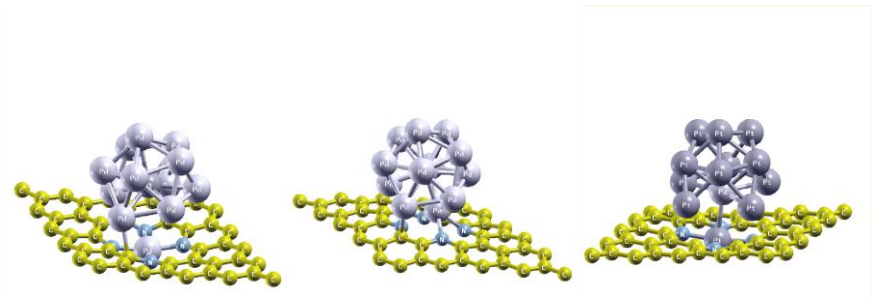


Figura 2 – Sustrato piridínico  $N_4V_2$  (contenido de nitrógeno del 8,3%) y un conjunto de  $Pd_{13}$ ,  $PdPt_{12}$  y  $PtPd_{12}$  como catalizadores.

Durante la optimización se observó deformación del sustrato así como migración de átomos metálicos a la vacancia correspondiente al defecto  $N_4V_2$ .

## 5 REFERENCIAS

- [1]. Rangel E, Sansores E, Vallejo E, Hernández-Hernández A, LópezPérez PA (2016) Study of the interplay between N-graphene defects and small Pd clusters for enhanced hydrogen storage via a spill-over mechanism. *Phys Chem Chem Phys* 18(48):33158– 33170. <https://doi.org/10.1039/C6CP06497C> 2.

- [2]. López MJ, Blanco-Rey M, Juaristi JI, Alducin M, Alonso JA (2017) Manipulating the magnetic moment of palladium clusters by adsorption and dissociation of molecular hydrogen. *J Phys Chem C* 121(38):20756–20762. <https://doi.org/10.1021/acs.jpcc.7b03996>.
- [3]. Gerber IC, Serp P (2020) A theory/experience description of support effects in carbon-supported catalysts. *Chem Rev* 120(2):1250–1349. <https://doi.org/10.1021/acs.chemrev.9b00209>
- [4]. Rangel E, Sansores E (2014) Theoretical study of hydrogen adsorption on nitrogen doped graphene decorated with palladium clusters. *Int J Hydrogen Energy* 39(12):6558–6566. <https://doi.org/10.1016/j.ijhydene.2014.02.062>
- [5]. P. Giannozzi, S. Baroni, N. Bonini, M. Calandra and et.al. *J. Phys.:Condens. Matter*, 2009,21, 395502



# MANTENIMIENTO PREDICTIVO IA- OPTIMIZADO PARA INDUSTRIA 4.0

## **Carlos I. Ramírez Bello**

*Forte Innovation Consulting  
S.A.P.I de C.V.*

*Plaza de la paz 102, Suite  
605 en piso 6 del G100 Col.  
Guanajuato Puerto Interior  
Silao, Guanajuato. C.P  
36275.*

*Email:  
ramirezbelloci@gmail.com  
web page:  
https://forteinnovation.mx*

## **Gabriel Ramos Zacarias**

*Forte Innovation Consulting  
S.A.P.I de C.V.*

*Plaza de la paz 102, Suite  
605 en piso 6 del G100 Col.  
Guanajuato Puerto Interior  
Silao, Guanajuato. C.P  
36275.*

*Email:  
iamramosg@gmail.com  
web page:  
https://forteinnovation.mx*

## **Takawira Joseph Mumanga**

*Forte Innovation Consulting S.A.P.I  
de C.V.*

*Plaza de la paz 102, Suite 605 en piso  
6 del G100 Col. Guanajuato Puerto  
Interior Silao, Guanajuato. C.P  
36275.*

*Email:  
[joseph.mumanga@forteinnovation.mx](mailto:joseph.mumanga@forteinnovation.mx)  
web page: <https://forteinnovation.mx>*

**Resumen.** El proyecto "Mantenimiento Predictivo AI-Optimizado para Industria 4.0" busca mejorar la eficiencia operativa en maquinaria industrial mediante un sistema avanzado de mantenimiento predictivo basado en inteligencia artificial (IA) y Machine Learning (ML). Desarrollado por Forte Innovation Consulting en el Laboratorio de Supercómputo del Bajío (Lab-SB) de CIMAT, el proyecto integra datos en tiempo real y modelos de red neuronal densa (MLP) para anticipar fallos y optimizar el mantenimiento. Se entrenaron modelos utilizando técnicas de batch learning y fine-tuning, evaluando el impacto del tamaño del dataset y la configuración de los modelos en los requerimientos de computación para su implementación. Para Failure Based Predictive Maintenance (FBPM), se utilizó un dataset público (AI4I), mientras que para Condition Based Predictive Maintenance (CBPM), se generó un dataset propio a partir de un prototipo.

**Palabras clave:** mantenimiento predictivo, inteligencia artificial, machine learning, deep learning, Industria 4.0, condition-monitoring, failure-prediction, supercómputo, paralelismo.

## INTRODUCCIÓN

El mantenimiento predictivo industrial surgió como una evolución de las prácticas de mantenimiento preventivo y correctivo, con el objetivo de anticipar fallos antes de que ocurran, minimizando tiempos de inactividad y optimizando los costos de operación [1]. Este enfoque es posible gracias al avance de la tecnología en sensores, análisis de datos y modelado predictivo, con el monitoreo de vibraciones y otros parámetros físicos en maquinaria [2]. Con el tiempo, el mantenimiento predictivo ha integrado tecnologías avanzadas como la inteligencia artificial y el aprendizaje automático, permitiendo un análisis más preciso y en tiempo real del estado de los equipos, y convirtiéndose en una práctica esencial en la gestión de activos industriales.

En el mantenimiento predictivo, "Condition-Monitoring" y "Failure-Prediction" son dos enfoques clave. Utiliza sensores y técnicas de análisis de datos para evaluar el rendimiento y las condiciones de funcionamiento en tiempo real. El "Condition-Monitoring" se enfoca en monitorear continuamente los parámetros que representan estado de una máquina para detectar anomalías que puedan indicar un problema futuro [3]. Por otro lado, "Failure-Prediction" se concentra en monitorear continuamente la correlación entre los parámetros que representan el estado de una máquina para predecir cuándo ocurrirá un fallo en ella antes de que suceda, utilizando algoritmos de aprendizaje automático y análisis predictivo [4]. Este enfoque se basa en patrones históricos y datos operativos para anticipar fallos, permitiendo intervenciones proactivas [5]. Ambos métodos son complementarios y esenciales para un mantenimiento predictivo efectivo [1].

Este informe muestra el enfoque que siguió el equipo del Centro de Innovación y Desarrollo (Centro I+D) de Forte Innovation Consulting usando entornos del Laboratorio de Supercómputo del Bajío (Lab-SB) de CIMAT para desarrollar y madurar su sistema de mantenimiento predictivo para industria. La meta es implementar modelos de IA de monitoreo capaces de analizar la información obtenida de sensores de máquinas industriales para detectar posibles fallas e informar cuándo es necesario realizar mantenimiento. Los modelos entrenados deben predecir valores y condiciones futuras fuera del rango normal, permitiendo un mantenimiento más eficiente, eficaz y pertinente [6].

También fue importante la evaluación del comportamiento y eficacia de los modelos creados en ejecuciones simultáneas, esto con el objetivo de aproximar las necesidades de la industria real de supervisar múltiples factores de forma simultánea e incluso con múltiples máquinas al mismo tiempo [7]. Este enfoque, conocido como paralelismo, consiste en la ejecución

simultánea de múltiples procesos o tareas para mejorar la eficiencia y reducir el tiempo de procesamiento. Para ello se diseñaron pruebas con paralelismo midiendo las limitaciones que ofrecían varias opciones computacionales.

## Relevancia e importancia

La implementación de mantenimiento predictivo ofrece numerosas ventajas para la industria:

- **Reducción de tiempos de inactividad:** Al predecir y prevenir fallos antes de que ocurran, se minimizan los periodos de inactividad no planificados.
- **Optimización de recursos:** Permite planificar los mantenimientos de manera más eficiente, optimizando el uso de recursos humanos y materiales.
- **Incremento en la vida útil de los equipos:** Un mantenimiento oportuno y preciso puede prolongar la vida útil de la maquinaria, reduciendo la necesidad de reemplazos frecuentes.

## OBJETIVOS

### Objetivo General

El objetivo principal del proyecto es desarrollar, evaluar y optimizar la implementación del sistema de mantenimiento predictivo usando entornos del Laboratorio de Supercómputo del Bajío (Lab-SB) de CIMAT. La implementación de este proyecto permitirá predecir fallas y condiciones operativas fuera del rango normal, optimizando así la eficiencia y la eficacia del mantenimiento de la maquinaria industrial.

- **Desarrollo de los Modelos:** Desarrollar modelos de predicción de fallos utilizando datasets específicos y con diferentes configuraciones de datapoints. Y diseñar modelos de predicción de condiciones.
- **Evaluación del Rendimiento y Precisión:** Evaluar el rendimiento y precisión de los modelos en diferentes entornos y con distintos parámetros de entrenamiento, midiendo métricas como precisión, F1, MAE, MSE y RMSE [8].
- **Optimización de Recursos para Implementación 'On Premise':** Optimizar parámetros y calcular los recursos necesarios para la implementación 'on premise' del proyecto en el entorno industrial, utilizando las métricas y requerimientos del modelo que demostró mejor precisión y eficiencia.
- **Evaluación de Rendimiento Computacional en ejecuciones en Paralelo:** Definir la carga de entrenamiento posible para el entorno

de Supercómputo y para otros entornos computacionales, teniendo un seguimiento de los parámetros de precisión mientras se aumenta la demanda de modelos en paralelo de forma progresiva [9].

## ALCANCE

### Alcance del Proyecto

Este proyecto cubrirá los siguientes aspectos:

- **Desarrollo y prueba de modelos:** Se han desarrollado modelos base de predicción de fallos utilizando Batch-Learning. Estos modelos han sido entrenados y validados utilizando el Datasets público de “Machine Failure” de AI4I. También se ha escalado dicho dataset para evaluarla eficacia del entrenamiento con un mayor número de datapoints. Además, se ha creado un dataset propio de un prototipo que contiene sensores, placas de Arduino y robots para llevar a cabo pruebas de entrenamiento de modelos para el “Condition Monitoring”.

### Limitaciones

El proyecto enfrenta las siguientes limitaciones:

- **Datos disponibles:** El proyecto no está diseñado para un tipo de maquinaria en específico, por lo cual el Dataset deberá adaptarse para la implementación industrial para cada máquina, y considerando los sensores disponibles.

## METODOLOGÍA

### Métodos de recopilación de datos

#### Fuentes de datos:

- **Datasets sintéticos y reales:** Se utilizó el 'AI4I 2020 Predictive Maintenance Dataset' con 10 mil Datapoints originales como fuente de datos de predicción de fallos. Para escalar los datos, se empleó la herramienta de Mostly AI, generando un mayor volumen de datos sintéticos (50 mil y 100 mil).
- **Uso de datos en tiempo real:** Se utilizó un prototipo con sensores para la lectura en tiempo real de factores ambiente y el almacenamiento de dichos datos históricos para el entrenamiento de los modelos de predicción de condiciones.

#### Distribución de datos:

- Los datos se distribuyeron en un 80% para el entrenamiento de los modelos, un 10% para la validación y un 10% para las pruebas.

## Herramientas y técnicas utilizadas

### Herramientas:

- Lenguajes de programación y bibliotecas: Se utilizaron bibliotecas de Python, como TensorFlow y Keras, para el desarrollo y entrenamiento de modelos de red neuronal densa. Para la visualización de datos, se emplearon herramientas de graficación de la librería de Python Matplotlib, así como una interfaz WEB con Chart.js y otra en P.
- Bases de datos y sistemas operativos: Los datos se almacenaron en archivos con extensión csv y se utilizó entornos de Linux y Windows para desarrollar, entrenar y evaluar los modelos de machine learning.

### Técnicas de machine learning:

- Batch Learning: Se empleó Batch Learning para el entrenamiento inicial de los modelos y se planifica el uso de online learning para el ajuste en tiempo real de los modelos con datos operativos.
- Modelos de red neuronal densa: Se utilizaron redes neuronales densas para la predicción de fallos. Esta red neuronal está diseñada para problemas de clasificación binaria, en este caso, para predecir la columna "Machine failure" basada en las características de entrada proporcionadas.

### Proceso paso a paso:

El proceso de entrenamiento de los modelos contó con pasos sucesivos concretos. Principalmente se enfocaron en generar datos sintéticos, cargar y preparar datos reales y sintéticos, entrenar un modelo de red neuronal densa para predecir fallas en máquinas, monitorear el uso de memoria durante el entrenamiento, evaluar el rendimiento del modelo con varias configuraciones, y finalmente visualizar y guardar los resultados obtenidos. En más detalle se siguió la siguiente estructura:

1. Generación de Datos Sintéticos: Uso de dataset de I4AI y escalado con Mostly AI.
2. Preparativos: Importación de bibliotecas para datos, modelos y monitoreo de memoria.
3. Monitoreo de Memoria: Función de registro de uso de memoria en segundo plano.
4. Carga y Preparación de Datos: Lectura de CSV, configuración de índice y conversión de texto a números.
5. División de Datos: División en conjuntos de entrenamiento (80%), prueba (10%) y validación (10%).



6. Entrenamiento de Modelos: Entrenamiento de modelos de red neuronal densa con diferentes configuraciones de datasets y epochs.
7. Creación del Modelo: Construcción y compilación de un modelo de red neuronal con Keras.
8. Entrenamiento del Modelo: Entrenamiento y validación del modelo, con monitoreo de memoria.
9. Validación y Pruebas: Procesamiento de datos, entrenamiento y guardado del modelo.
10. Evaluación del Rendimiento: Evaluación con métricas y variación de epochs y volúmenes de datasets.
11. Evaluación del Modelo: Cálculo de precisión, puntaje F1 y creación de gráficos comparativos.
12. Visualización de Resultados: Creación y guardado de gráficos de comparación y precisión.

## HALLAZGOS

### Detalles técnicos

La computadora de CIMAT cuenta con 48GB de RAM distribuidos en dos GPUs de 24GB y la PC Personal cuenta con 16GB de RAM.

### Matriz de escenarios

Esta tabla presenta un resumen comparativo de los resultados obtenidos al entrenar modelos de redes neuronales densas en diferentes plataformas, con variaciones en la cantidad de datos (datapoints) y épocas (epochs). Se evalúan las métricas clave de precisión (Accuracy), puntaje F1 (F1 Score), tiempo de procesamiento (Tiempo en segundos), y el uso de recursos computacionales (Procesamiento en MB/s) en entornos como CIMAT Linux, Computadora Personal, y Google Colab. La tabla destaca las diferencias en rendimiento y eficiencia según la configuración del modelo y el entorno de ejecución, proporcionando una visión clara de cómo estos factores afectan el desempeño del sistema de mantenimiento predictivo.

ID	Plataforma	Datapoints	Epochs	Accuracy	F1 Score	Tiempo (s)	Procesamiento (MB/s)
1	CIMAT Linux	10,000	10	98.10%	0.63	10.99	2030,65
2	CIMAT Linux	50,000	10	96.86%	0.08	50.29	2005,27
3	CIMAT Linux	100,000	10	96,86%	0.14	104.05	2071,48
4	CIMAT Linux	10,000	100	98,1% - 98,50%	0,69 - 0,75	99.41	1925,26

5	CIMAT Linux	50,000	100	96.68%	0.23	488.44	1976.89
6	CIMAT Linux	100,000	100	96,71%	0.13	976.22	1954,56
7	<b>Computadora Personal</b>	<b>10,000</b>	<b>10</b>	<b>97.70%</b>	<b>0.61</b>	<b>5.45</b>	<b>1048.84</b>
8	Computadora Personal	50,000	10	96.90%	0.16	21.13	1070.36
9	Computadora Personal	100,000	10	96.92%	0.11	40.92	1101.99
10	<b>Computadora Personal</b>	<b>10,000</b>	<b>100</b>	<b>98.30%</b>	<b>0.71</b>	<b>106.19</b>	<b>1067.96</b>
11	Computadora Personal	50,000	100	96.76%	0.29	222.33	1076.88
12	Computadora Personal	100,000	100	96.77%	0.13	421.69	1102.16
13	<b>Google Colab Ubuntu Linux</b>	<b>10,000</b>	<b>10</b>	<b>97.80%</b>	<b>0.65</b>	<b>21.74</b>	<b>1586.03</b>
14	Google Colab Ubuntu Linux	50,000	10	96.86%	0.09	83.33	1538.22
15	Google Colab Ubuntu Linux	100,000	10	96.94%	0.08	143.84	1575.42
16	<b>Google Colab Ubuntu Linux</b>	<b>10,000</b>	<b>100</b>	<b>98.70%</b>	<b>0.78</b>	<b>105.07</b>	<b>1537.92</b>
17	Google Colab Ubuntu Linux	50,000	100	96.82%	0.26	563.50	1524.39
18	Google Colab Ubuntu Linux	100,000	100	96.77%	0.19	1043.36	1552.21

Tabla 1 : Tabulación de la matriz de escenarios y sus resultados para ejecuciones de variantes de modelos sin paralelismo

## Interpretación de resultados graficados

Este gráfico compara el tiempo de entrenamiento en segundos de un modelo de red neuronal densa utilizando 10,000 datos en diferentes plataformas: CIMAT, Computadora Personal y Google Colab. Se presentan los resultados para dos configuraciones: 10 épocas y 100 épocas.

Entrenamiento con 10,000 Datos

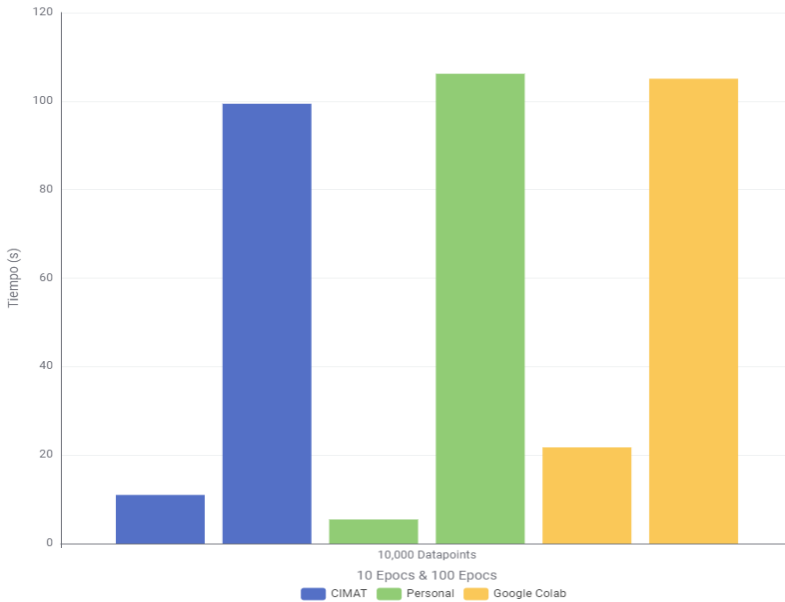


Figura 1 – Interpretación de resultados de rendimiento de los escenarios ejecutados

El eje X muestra los casos de ejecución, siendo estos la computadora en la que se realizó la ejecución, el número de epochs de la prueba (A la izquierda 10 epochs y a derecha 100 para cada computadora) y también se detalla que todos los modelos se entrenan usando el Dataset de 10 mil datapoints, el cual es el original obtenido de AI41 sin escalar. Esto debido a que el escalamiento usando Mostly AI alteró las relaciones entre los datos y por tanto perjudicó la eficacia de los modelos que se entrenaron con ellos.

El entrenamiento de cada modelo fue realizado en un contenedor individual usando Docker y el registro de consumo de memoria se almacena en dicho contenedor. Del mismo modo, usando Docker-Compose, las ejecuciones de modelos en paralelo ejecutaron de forma simultánea una cantidad predefinida de contenedores y en cada uno se registra la memoria consumida de forma global durante su periodo de ejecución.

**NOTA:** La exclusión de 1,000 epochs en la matriz de escenarios se debe a la necesidad de optimizar el tiempo de entrenamiento y los recursos. Entrenar con 1,000 epochs puede ser impráctico en términos de tiempo y recursos, especialmente en plataformas con menor capacidad de procesamiento. Además, después de cierto punto, los beneficios adicionales

de más epochs son mínimos. Se han realizado pruebas exploratorias con 1,000 epochs únicamente en la supercomputadora del CIMAT.

### Elección del modelo con 10 mil datapoints y 100 epochs

Como se observa en los resultados el modelo con 10 mil datapoints y 100 epochs tuvo el mejor rendimiento con respecto a las métricas F1, por la gran diferencia de precisión se escogieron los parámetros de este modelo para ejecutar las pruebas de paralelismo. Es importante mencionar que el dataset con 10 mil datapoints es el único que no se escaló por lo que los datos son fieles al dataset I4AI sin alteraciones, derivado de los resultados se considera que hay una alta probabilidad de que la precisión deficiente obtenida por el resto de los volúmenes de datasets se deba al escalamiento.

### Matriz de pruebas base con paralelismo

La siguiente tabla presenta un análisis comparativo de las pruebas realizadas para evaluar el rendimiento de modelos de red neuronal densa bajo condiciones de paralelismo. Estas pruebas fueron ejecutadas en diferentes entornos computacionales (CIMAT Linux y Computadora Personal) utilizando una configuración estándar de 10,000 datos y 100 épocas.

ID	Plataforma	Datapo ints	Epo chs	Paralel ismo	Accuracy	F1 Score	Tie mpo (s)	Procesa miento (MB/s)
1	CIMAT Linux	10,000	100	5	98.20%	0.69	85	3320
2	CIMAT Linux	10,000	100	10	98.20%	0.69	63	4744
3	CIMAT Linux	10,000	100	15	98.20%	0.69	81	6367
4	CIMAT Linux	10,000	100	30	98.20%	0.69	120	11 079
5	CIMAT Linux	10,000	100	50	98.20%	0.69	174	18 856
6	CIMAT Linux	10,000	100	70	98.20%	0.69	317	24 018
7	CIMAT Linux	10,000	100	100	98.20%	0.69	376	36 700
8	CIMAT Linux	10,000	100	150	98.20%	0.69	415	47 487
9	Computadora Personal Windows 11	10,000	100	5	98.70%	0.79	120	2383
10	Computadora Personal Windows 11	10,000	100	10	98.60%	0.77	220	3897
11	Computadora Personal Windows 11	10,000	100	15	98.50%	0.75	350	5521
12	Computadora Personal Windows 11	10,000	100	30	98.20%	0.69	690	7475

Tabla 2 : Tabulación de la matriz de escenarios y sus resultados para ejecuciones de variantes de modelos sin paralelismo

El límite alcanzado para el entrenamiento de modelos en paralelo por computadora se muestra en aquí:

<i>Computadora</i>	Límite de modelos en paralelo	Tiempo de ejecución	F1 score	Ram utilizado
<i>CIMAT</i>	150	415 s	0.69 en promedio	48/48
<i>PC PERSONAL</i>	30	733 s	0.75 en promedio	15.5/16

Tabla 3 : Tabulación de casos de paralelismo limite por computadora.

## CONCLUSIONES

Este estudio evaluó con éxito la efectividad y los requisitos computacionales de las redes neuronales densas para el mantenimiento predictivo basado en fallos en maquinaria industrial. Se demostró que los entornos computacionales avanzados, como la infraestructura del CIMAT, son cruciales para el rendimiento óptimo de estos modelos, especialmente en el manejo de grandes volúmenes de datos y entrenamientos prolongados.

Además, se identificaron parámetros de entrenamiento óptimos, como el uso de 10,000 puntos de datos con 100 épocas, que logran un equilibrio entre precisión y eficiencia computacional. Los modelos desarrollados mostraron alta precisión y robustez, lo que es esencial para implementaciones de sistemas de mantenimiento predictivo que reducen el tiempo de inactividad no planificado y mejorar la eficiencia operativa en entornos industriales.

Estos hallazgos subrayan la importancia de alinear los recursos computacionales con las demandas del entrenamiento para maximizar la efectividad de las operaciones de mantenimiento predictivo. La comprensión de estos factores permite a las industrias hacer inversiones informadas en infraestructura, asegurando la escalabilidad y flexibilidad necesarias para adaptar los sistemas de mantenimiento predictivo a diversos contextos industriales.

- [1] D. Ženíšek, J. Bašl, and P. Poór, “Historical Overview of Maintenance Management Strategies: Development from Breakdown Maintenance to Predictive Maintenance in Accordance with Four Industrial Revolutions.” [Online]. Available: <https://www.researchgate.net/publication/335444202>
- [2] M. Pech, J. Vrchota, and J. Bednář, “Predictive maintenance and intelligent sensors in smart factory: Review,” Feb. 02, 2021, MDPI AG. doi: 10.3390/s21041470.

- [3] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mech Syst Signal Process*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006, doi: 10.1016/J.YMSSP.2005.09.012.
- [4] J. Leukel, J. González, and M. Riekert, “Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review,” Oct. 01, 2021, Elsevier B.V. doi: 10.1016/j.jmsy.2021.08.012.
- [5] F. Alves et al., “Deployment of a Smart and Predictive Maintenance System in an Industrial Case Study,” in *2020 IEEE 29th International Symposium on Industrial Electronics (ISIE)*, IEEE, Jun. 2020, pp. 493–498. doi: 10.1109/ISIE45063.2020.9152441.
- [6] P. Kamat and R. Sugandhi, “Anomaly Detection for Predictive Maintenance in Industry 4.0- A survey,” *E3S Web of Conferences*, vol. 170, p. 02007, May 2020, doi: 10.1051/e3sconf/202017002007.
- [7] P. Jawandhiya, “Hardware Design for Machine Learning,” *International Journal of Artificial Intelligence & Applications*, vol. 9, no. 1, pp. 63–84, Jan. 2018, doi: 10.5121/ijai.2018.9105.
- [8] G. Varoquaux and O. Colliot, “Evaluating Machine Learning Models and Their Diagnostic Value,” 2023, pp. 601–630. doi: 10.1007/978-1-0716-3195-9\_20.
- [9] S. Lee, J. Kyu Kim, X. Zheng, Q. Ho, G. A. Gibson, and E. P. Xing, “On Model Parallelization and Scheduling Strategies for Distributed Machine Learning.”



# EVALUACIÓN DE ESTRATEGIAS DE INGENIERÍA DE PROMTS PARA MODELO DE LENGUAJE A GRAN ESCALA (LLMS) EN EL PROCESAMIENTO DEL LENGUAJE NATURAL ACADÉMICO

**Guadalupe H. Rangel-  
Sánchez, Emma Sánchez-  
Barragán, Liliana Fausto-  
Castro, Aideé Hernández-  
López**

*Universidad del SABES (UNIDEG)  
Blvd. Guanajuato #1615, Real Providencia.  
C.P. 37234, León, Gto. México.  
Email: [aidee.hernandezl@sabes.edu.mx](mailto:aidee.hernandezl@sabes.edu.mx), web  
page: <https://sabes.edu.mx/portal/>*

**Resumen.** Los modelos de lenguaje a gran escala (LLMs) han evolucionado el procesamiento del lenguaje natural, mostrando capacidades relevantes en diversas tareas lingüísticas. Sin embargo, su implementación en el ámbito académico enfrenta desafíos en cuanto a precisión y pertinencia. Este estudio, es fruto de la colaboración entre la Universidad del SABES y el CIMAT en Guanajuato y analiza la eficiencia de estrategias de diseño de prompts en los LLMs al procesar textos académicos. La finalidad es identificar prácticas óptimas para mejorar la precisión y relevancia de las respuestas generadas en contextos académicos. La presente investigación, respaldada por IDEA Gto., utiliza el laboratorio de Supercómputo de CIMAT, permitiendo experimentos a gran escala, este enfoque es crucial dado el creciente uso de los LLMs en entornos educativos y de investigación, donde la exactitud de la información es fundamental.

**Palabras clave:** Modelos de lenguaje a gran escala, LLM, diseño de prompts, lenguaje natural, textos académicos.



## INTRODUCCIÓN

### Ingeniería de Prompts en LLMs

Los LLMs son sistemas computacionales que utilizan arquitecturas de redes neuronales profundas, como las redes neuronales transformadoras (Transformers networks)<sup>1</sup>, para procesar y generar lenguaje humano. Estos modelos, como GPT-3<sup>2</sup> han revolucionado la inteligencia artificial aplicada al lenguaje, permitiendo tareas desde la traducción automática hasta la generación de texto creativo.

#### Ingeniería de Prompts.

La ingeniería de prompts es el proceso de diseñar y crear instrucciones o indicaciones que provocan las respuestas deseadas de los LLMs. Las indicaciones pueden clasificarse en diferentes tipos según su estructura, función y complejidad<sup>3</sup>. La ingeniería de prompts es crucial para guiar la salida del modelo hacia la intención del usuario, minimizando la necesidad de ajustes finos o de grandes cantidades de datos de entrenamiento adicionales<sup>4</sup>. Se pueden clasificar a las indicaciones (prompts) en 7 tipos<sup>3</sup>, como se ilustra en la Figura 1

Tipos de Prompts para LLMs

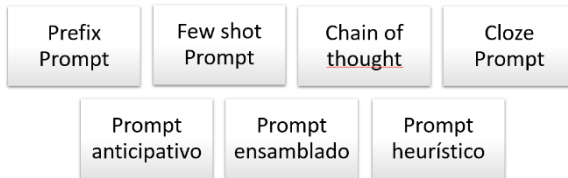


Figura 1- Tipos de prompts<sup>3</sup>

La ingeniería de prompts se ha convertido en una práctica clave para optimizar la interacción entre los usuarios (en este caso, estudiantes y profesores) y los LLMs. Al diseñar prompts específicos y contextualizados, se puede guiar a los modelos hacia la extracción de información más precisa y relevante para las necesidades educativas. Por ejemplo, prompts que incluyen terminología y contextos académicos específicos pueden mejorar la precisión de las respuestas proporcionadas por los LLMs<sup>5</sup>.

## Estrategias de Ingeniería de Prompts

### *Prompt Design*

El diseño de prompts implica la creación de entradas claras y específicas a las LLMs que guíen el modelo hacia la respuesta deseada. Por lo que se plantea el cuestionamiento de cómo el formato, la longitud y la especificidad del prompt pueden influir en el rendimiento del modelo<sup>6</sup>

### *Prompt Tuning*

En esta estrategia el enfoque radica en ajustar dinámicamente los prompts durante el uso del modelo para mejorar la precisión y la relevancia de las respuestas. Se han explorado técnicas de prompt tuning en aplicaciones de pregunta y respuesta (Question and Answer)<sup>7</sup>.

### *Prompt Augmentation*

La ampliación de prompts incluye la adición de contexto o ejemplos a los prompts para mejorar la comprensión del modelo. Se ha utilizado esta técnica para mejorar la generación de texto en tareas específicas<sup>5</sup>.

### *Evaluación de la Efectividad de los Prompts*

La evaluación de la efectividad de los prompts es esencial para garantizar que los LLMs proporcionen resultados útiles y confiables. Métricas como la precisión, la fluidez del lenguaje y la relevancia contextual son cruciales<sup>8</sup>. Algunos estudios<sup>9</sup> abordan cómo la evaluación puede ser automatizada y adaptada a diferentes contextos de aplicación. La efectividad de los prompts propuestos en<sup>3</sup> se mide evaluando el rendimiento de los modelos de lenguaje en tareas de procesamiento de lenguaje natural utilizando algunas métricas como precisión y exactitud.

## La extracción de información (EI)

Es un componente crucial en el ámbito educativo, donde la precisión y la relevancia de la información suministrada a los estudiantes pueden influir significativamente en su aprendizaje y comprensión de los temas académicos. En entornos universitarios, donde los estudiantes deben manejar grandes volúmenes de información y desarrollar habilidades de investigación y análisis crítico, la precisión en la EI es fundamental<sup>10</sup>.

Tradicionalmente, la EI en entornos educativos ha enfrentado retos importantes, como la ambigüedad del lenguaje natural, la heterogeneidad de los datos y la necesidad de procesar información en tiempo real. Estos retos han limitado la efectividad de las herramientas de EI, lo que a menudo resulta

en respuestas poco claras o inexactas, especialmente cuando se trata de contextos académicos específicos y términos técnicos<sup>11</sup>.

Con la aparición de los LLMs, como los que se listan a continuación:

GPT-4 (OpenAI), GPT-3 (OpenAI), LLaMA (Meta/Facebook), Gemini (Google), Mixtral 8x7B-Instruct-v0.1, BERT (Bidirectional Encoder Representations from Transformers) de Google, RoBERTa (A Robustly Optimized BERT Pretraining Approach) de Facebook, XLNet (Google/CMU), Megatron (NVIDIA), Turing-NLG (Microsoft), GPT-Neo (EleutherAI), Gopher (DeepMind), Claude, Deep Seek.

Se ha abierto un nuevo horizonte en la capacidad de procesar y generar lenguaje natural con una gran precisión y fluidez. Estos modelos, entrenados en vastos corpus de texto, han demostrado ser capaces de manejar una amplia gama de tareas de EI, desde la comprensión del contexto hasta la generación de resúmenes y respuestas a preguntas específicas<sup>12</sup>.

La evaluación de la precisión en la EI utilizando LLMs e ingeniería de prompts es un área activa de investigación. Se requiere el desarrollo de métricas y benchmarks específicos para el ámbito educativo que consideren no solo la exactitud de las respuestas, sino también su relevancia y adecuación al contexto académico<sup>13</sup>.

La precisión en la EI en ámbitos universitarios y educativos utilizando ingeniería de prompts y LLMs es un campo de investigación en crecimiento que aborda importantes retos y oportunidades en la mejora del aprendizaje y la enseñanza mediante la utilización de tecnologías de inteligencia artificial.

Se han encontrado aplicaciones de la ingeniería de prompts a diversas áreas del conocimiento, por ejemplo, en el diagnóstico médico<sup>3</sup> y se han observado avances importantes en esa dirección, sin embargo, existen áreas de oportunidad muy relevantes para investigar, en especial en el área académica. A pesar de su potencial, la investigación en este campo aún es limitada, y no está claro cuáles de estas estrategias son más efectivas en el contexto académico, donde los textos pueden ser complejos y especializados<sup>1</sup>.

En esta investigación se busca abordar este problema mediante la evaluación de la eficiencia de diversas estrategias para el diseño de prompts en el procesamiento de textos académicos (específicamente en los apartados constituyentes de una tesis de licenciatura) por parte de los LLMs. La investigación tendrá como objetivo analizar la eficiencia de las estrategias de diseño de prompts en el contexto de modelos de lenguaje de gran escala (LLMs) al procesar y analizar textos académicos en lenguaje natural, con el propósito de identificar las mejores prácticas que permitan mejorar la precisión y la generación de respuestas relevantes y útiles en el ámbito académico. Esto es crucial no solo para el avance de la tecnología del procesamiento del lenguaje natural (PLN) en el sector educativo y de investigación, sino también para fomentar que los recursos digitales sean cada

vez más accesibles y útiles para los estudiantes, investigadores y académicos. De tal manera que se pueda maximizar el potencial de los LLMs en el PLN, contribuyendo así a la mejora de la calidad y la accesibilidad de la información en el ámbito académico.

## MÉTODO

### Modelo metodológico

Para llevar a cabo la investigación se planteó el Método presentado en la Figura 2.

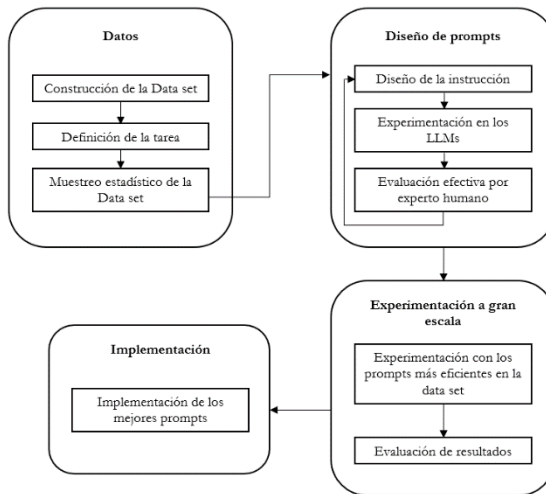


Figura 2- Método de investigación

La primera etapa mostrada en la Figura 2 se denomina “Datos”, en esta etapa, se desarrolla la construcción de la data set a analizar, que consiste en una base de datos integrada por la información contenida en las tesis presentadas por estudiantes de la Universidad del SABES durante los años 2022 y 2023. Una vez que el conjunto de datos esté estructurado, se determinarán las tareas específicas que se solicitarán a los distintos LLMs mediante la formulación de prompts estructurados. Finalmente, se llevará a cabo una selección aleatoria de la muestra estadística del conjunto de datos, con el propósito de obtener una muestra representativa en la cual se aplicarán y evaluarán diversas estrategias de diseño de prompts.

En la etapa denominada "Diseño de prompts", se inicia con la conformación de los prompts, en la cual se formulan una serie de instrucciones diversas y múltiples con el propósito de guiar a los LLMs en la realización de las tareas solicitadas. Posteriormente, se lleva a cabo una experimentación en la que se prueban los prompts diseñados en los distintos LLMs. Finalmente, se realiza una evaluación de la efectividad de los LLMs mediante la revisión de un experto humano, quien proporciona un juicio cualitativo y cuantitativo sobre el rendimiento de los modelos en función de los prompts aplicados, este proceso se realiza de manera iterativa, hasta que se encuentren los prompts que reporten las mejores eficiencias en la identificación de las tareas establecidas.

Posterior a la experimentación en la muestra representativa de la data set, se realiza en la etapa denominada "Experimentación a gran escala" que se refiere a la aplicación de los prompts mejor evaluados a toda la data set, mediante los recursos informáticos con los que cuenta el laboratorio de Supercómputo del CIMAT en Guanajuato. Posteriormente, durante la etapa de "Implementación", se elabora un catálogo detallado que incluye todos los prompts identificados y evaluados en las fases anteriores del proceso investigativo.

### **Instrumentos de evaluación**

Para llevar a cabo el proyecto que involucra el análisis de la eficiencia de las estrategias de ingeniería de prompts en los LLMs al procesar y analizar textos académicos en lenguaje natural, específicamente en la evaluación de los objetivos generales de tesis de licenciatura, se recurrió a los siguientes análisis **Análisis de la eficiencia.** Para encontrar el porcentaje de efectividad de la instrucción y la salida que provee el LLM en el procesamiento y análisis de textos académicos se desarrolló una matriz comparativa entre la entrega del LLM y la verificación del experto humano, cuando se encuentran diferencias en la opinión y la entrega se coloca un número (1) y cuando ambas entregas coinciden, entonces se coloca un número (0).

**Análisis comparativo.** Se realizó un análisis comparativo entre diferentes estrategias de prompts y LLMs. Esto permite la identificación de qué estrategias son más efectivas en diferentes contextos y para diferentes tipos de textos académicos.

## **RESULTADOS**

En el presente estudio, se recopiló una Data set que abarca las tesis desarrolladas entre los años 2022 y 2023 en las distintas facultades de la Universidad del SABES. Esta Data set incluye los títulos de las tesis, sus objetivos generales y específicos, así como las hipótesis propuestas. Para la

fase experimental, se realizó un muestreo aleatorio de las tesis contenidas en la Data set, seleccionando únicamente los objetivos generales con el propósito de analizar sus componentes estructurales.

Posteriormente, se definieron las tareas específicas que se solicitaron a los LLMs, las cuales se detallan a continuación:

- a. Identificación del verbo principal del objetivo general en modo infinitivo.
- b. Identificación del tema de investigación.
- c. Identificación del alcance o contexto de la investigación.
- d. Identificación del propósito o finalidad de la investigación.
- e. Encontrar si el objetivo tiene congruencia gramatical.
- f. Evaluación general del objetivo de la tesis.

Se sometieron a prueba diferentes tipos de prompts, incluyendo:

- a. Zero shot
- b. Chain of thought (CoT)
- c. Few shot-CoT
- d. Árbol de pensamiento

Estos prompts se evaluaron con los siguientes LLMs:

- a. Deep Seek
- b. Llama 3.1-70B Instruct
- c. Mixtral 8x7B Instruct-v0.1
- d. GPT-3.5-turbo-0613

Los resultados de la efectividad alcanzada en las tareas encomendadas a los LLMs, empleados en este estudio, se detallan en la *Tabla 1*.

Tarea y LLMs	Zero Shot		CoT		Few Shot-Cot		Árbol de pensamiento	
<i>Identificación del verbo principal en modo infinitivo</i>	a	84%	a	88%	a	86%	<i>na</i>	
	b	88%	b	92%	b	90%		
	c	47%	c	84%	c	90%		
	d	84%	d	86%	d	94%		
<i>Identificación del tema de investigación</i>	a	92%	a	96%	a	96%	<i>na</i>	
	b	90%	b	63%	b	80%		
	c	90%	c	82%	c	90%		
	d	88%	d	90%	d	92%		
<i>Identificación del alcance o contexto</i>	a	80%	a	82%	a	82%	<i>na</i>	
	b	80%	b	76%	b	78%		
	c	75%	c	74%	c	77%		
	d	69%	d	53%	d	78%		
<i>Identificación del propósito o finalidad</i>	a	76%	a	76%	a	67%	<i>na</i>	
	b	69%	b	96%	b	73%		
	c	71%	c	65%	c	73%		
	d	65%	d	69%	d	65%		
<i>Identificación de la congruencia gramatical</i>	a	18%	a	22%	a	27%	<i>na</i>	
	b	63%	b	49%	b	33%		
	c	43%	c	31%	c	33%		
	d	29%	d	16%	d	35%		
<i>Evaluación general del objetivo</i>	<i>na</i>		<i>na</i>		<i>na</i>		a	47%
	<i>na</i>		<i>na</i>		<i>na</i>		b	71%
	<i>na</i>		<i>na</i>		<i>na</i>		c	61%
	<i>na</i>		<i>na</i>		<i>na</i>		d	63%

a. *Deep seek*, b. *Meta-Llama 3.1-70B-Instruct* c. *Mistral-8x7B-Instruct-v0.1*,  
d. *GPT-3.5-turbo-0613*. na (no aplica para este caso)

Tabla 1: Resumen de la efectividad lograda en cada tarea por las LLMs.

En la *Tabla 2* se muestran los prompts que tuvieron mayor efectividad al probarlos en los diferentes LLMs.

Tarea	Prompt			
	Zero Shot	CoT	Few shot-CoT	Árbol de pensamiento
<i>Identificación del verbo principal en modo infinitivo</i>	Ver Anexo 1. Sección 1	Ver Anexo 2. Sección 1	Ver Anexo 3. Sección 1	<i>na</i>
<i>Identificación del tema de investigación</i>	Ver Anexo 1. Sección 2	Ver Anexo 2. Sección 2	Ver Anexo 3. Sección 2	<i>na</i>
<i>Identificación del alcance o contexto</i>	Ver Anexo 1. Sección 3	Ver Anexo 2. Sección 3	Ver Anexo 3. Sección 3	<i>na</i>
<i>Identificación del propósito o finalidad</i>	Ver Anexo 1. Sección 4	Ver Anexo 2. Sección 4	Ver Anexo 3. Sección 4	<i>na</i>
<i>Identificación de la congruencia gramatical</i>	Ver Anexo 1. Sección 5	Ver Anexo 2. Sección 5	Ver Anexo 3. Sección 5	<i>na</i>
<i>Evaluación general del objetivo</i>	<i>na</i>	<i>na</i>	<i>na</i>	Ver Anexo 4

na (no aplica para este caso)

Tabla 2: Prompts que obtuvieron mayor efectividad en las tareas definidas

En la *Figura 3* se presentan de manera gráfica los resultados de la eficiencia lograda en el LLM Meta-Llama 3.1-70B-Instruct mediante los prompts diseñados para identificar los elementos que constituyen los objetivos generales en investigaciones, específicamente en tesis de licenciatura de la Universidad del SABES. Se evidencia un rendimiento mayor en tareas de menor complejidad. No obstante, ante tareas que implican un mayor nivel de dificultad, se registra una disminución significativa en la efectividad, llegando a un mínimo del 33%.



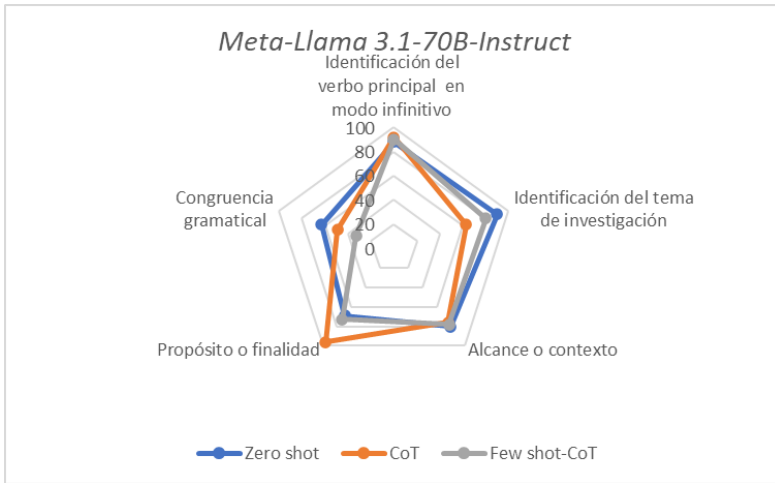


Figura 3- Eficiencia en la identificación de las tareas en el LLM Meta-Llama 3.1-70B-Instruct.

El análisis de las Figuras 4 y 6 revela patrones de eficiencia consistentes entre los diversos tipos de prompt para la mayoría de las tareas evaluadas, particularmente aquellas de menor complejidad. Sin embargo, se observa una notable excepción en la tarea de congruencia gramatical, donde la eficiencia experimenta una marcada disminución, alcanzando un valor mínimo del 16%.

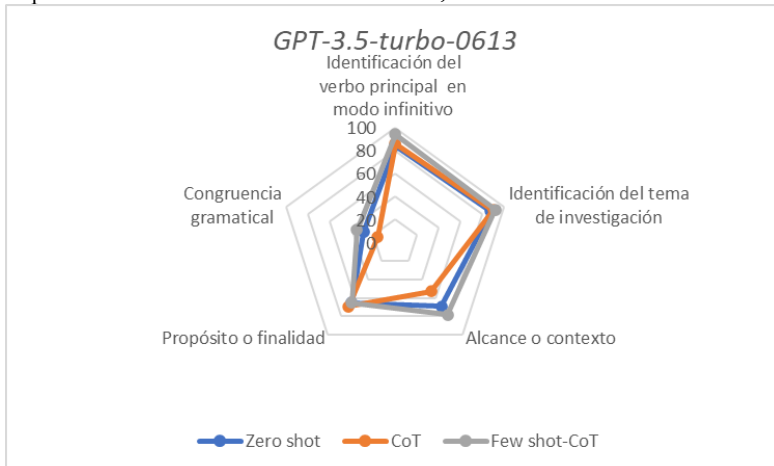


Figura 4- Eficiencia en la identificación de las tareas en el LLM GPT-3.5-turbo-0613.

Los resultados del análisis comparativo entre diversos Modelos de Lenguaje Grandes (LLMs, por sus siglas en inglés) revelan hallazgos significativos. Como se ilustra en la Figura 5, se observa una eficiencia del 47% en la tarea "Identificación del verbo principal en modo infinitivo" para uno de los

modelos evaluados. Este resultado es particularmente notable, ya que no se replica en los demás LLMs analizados. Adicionalmente, el modelo Mixtral-8x7B-Instruct-v0.1 exhibe una eficiencia del 31% en la evaluación de congruencia gramatical.

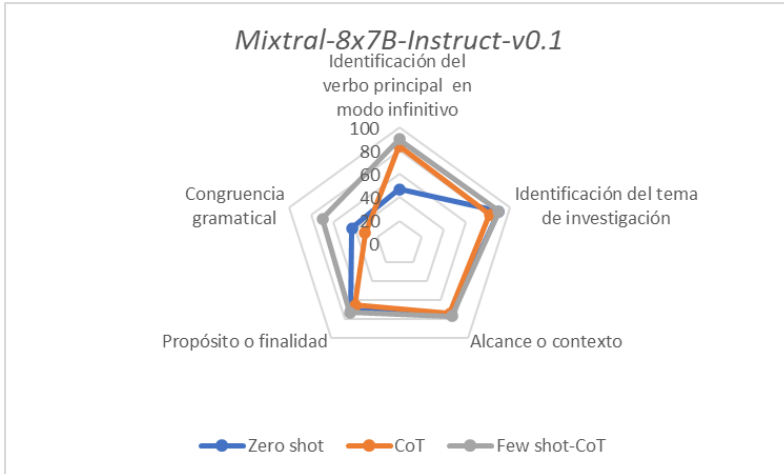


Figura 5- Eficiencia en la identificación de las tareas en el LLM Mixtral-8x7B-Instruct-v0.1.

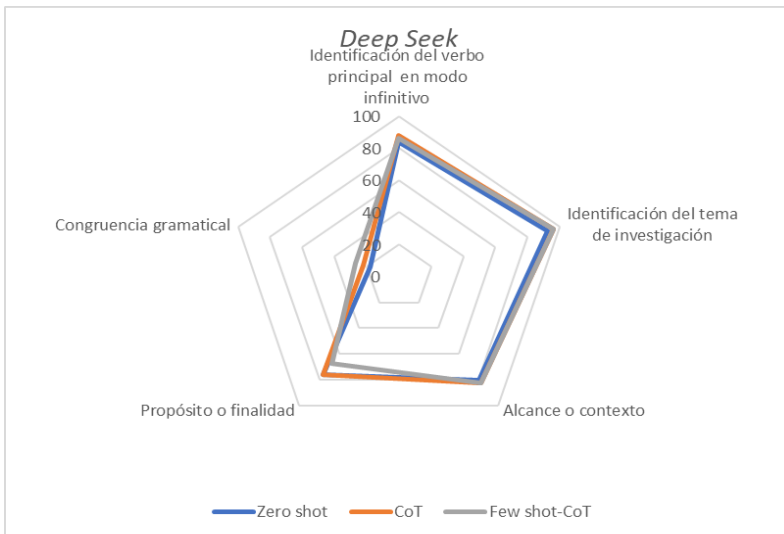
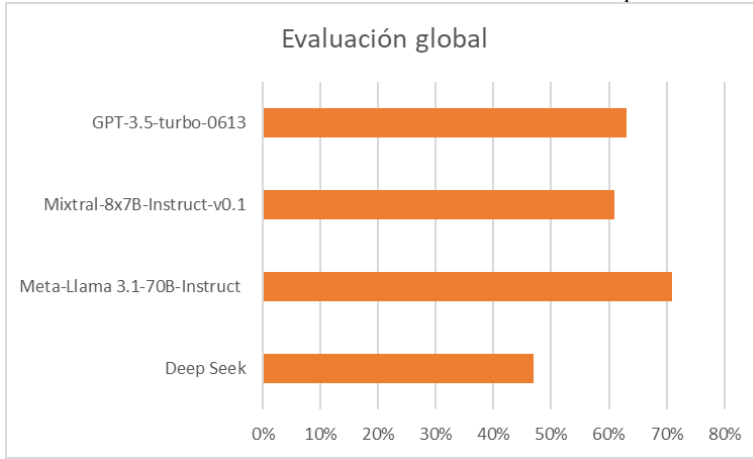


Figura 6- Eficiencia en la identificación de las tareas en el LLM Deep Seek

En la *Figura 7* se muestran los resultados de la tarea denominada "Evaluación global del objetivo general", la cual fue realizada mediante un prompt diseñado en forma de árbol de pensamiento. Los hallazgos indican que el modelo Meta-llama 3.1-70B-Instruct obtuvo la mayor eficiencia en la realización de esta tarea, alcanzando un 71% de desempeño.



*Figura 7- Evaluación global del Objetivo General de investigación mediante el prompt en árbol de pensamiento.*

Para la realización de la experimentación a gran escala, se empleó un nodo de Supercómputo equipado con dos unidades de procesamiento gráfico (GPU) TITAN RTX, cada una con 24 GB de memoria de video (VRAM) y 4.608 núcleos CUDA. Estas GPUs de alta gama demostraron ser fundamentales para el procesamiento eficiente de los modelos de lenguaje a gran escala (LLMs), los cuales demandan un considerable consumo de recursos computacionales.

Se utilizaron los siguientes LLMs cuantizados: gemma-7b, Llama-2-13b-chat-hf, Meta-Llama-3-70B-Instruct, Yi-1.5-34B-Chat, Mixtral-8x7B-Instruct-v0.1. Así mismo, se utilizaron los siguientes LLMs sin cuantizar: Llama-2-7b-chat-hf,

Meta-Llama-3-8B-Instruct, Meta-Llama-3.1-8B-Instruct, Mistral-7B-Instruct-v0.3. Los tiempos de procesamiento de los prompts en los diferentes LLMs variaron entre 73 y 250 minutos. El prompt utilizado en esta etapa de la investigación puede verse en el anexo 4.

Durante el curso de la investigación, se observó que algunos de los LLMs requerían una capacidad computacional que superaba las posibilidades de las GPUs TITAN RTX disponibles. Esto se debía a la enorme cantidad de parámetros que debían ser procesados y almacenados en la memoria de video (VRAM) de las GPUs. Por lo que se implementó la técnica de cuantización,

la cual permite reducir la precisión de los pesos del modelo, de 32 bits a 8 bits o incluso 4 bits. Esta disminución significativa de la precisión de los parámetros del modelo se tradujo en una reducción sustancial de la memoria requerida. Si bien esta reducción de la precisión introduce cierta pérdida de información, permitió que los modelos se ajustaran a las capacidades de las GPUs disponibles, manteniendo aun así un nivel de precisión razonable. De este modo, la aplicación de la técnica de cuantización resultó ser una estrategia fundamental para el procesamiento. No obstante, a pesar de la aplicación de la técnica de cuantización como estrategia fundamental para superar las limitaciones de recursos computacionales, los resultados obtenidos no fueron satisfactorios. Los LLMs no lograron discriminar eficientemente las tareas solicitadas, lo cual pone de manifiesto la necesidad de seguir explorando enfoques que permitan compatibilizar las exigencias computacionales de estos modelos con los recursos disponibles, sin comprometer sustancialmente su desempeño.

## DISCUSIÓN

En el trabajo presentado se ha llevado a cabo una exhaustiva evaluación de la eficiencia de las estrategias de diseño de prompts en los LLMs al procesar y analizar textos académicos en lenguaje natural. Los resultados obtenidos permiten establecer conclusiones relevantes en relación con los objetivos planteados:

Se identificaron diversas estrategias de ingeniería de prompt utilizadas en LLMs para el análisis de textos académicos. Estas estrategias incluyen el uso de prompts estructurados y prompts contextualizados, cada uno con su propósito específico en el procesamiento de información académica.

En la evaluación estructurada se realizaron ajustes finos al prompt de CoT con la ingeniería de indicaciones; la indicación del prompt detecta en el objetivo la estructura sintáctica con los elementos gramaticales incrementando así la eficiencia del prompt en la detección y logrando un léxico simplificado para la retroalimentación objetiva al universitario.

Las estrategias de ingeniería de prompt demostraron influencia en la precisión del procesamiento y análisis de textos académicos. Se encontró que los prompts diseñados con un enfoque más específico y contextualizado tienden a mejorar la precisión en la interpretación y generación de respuestas. Sin embargo, las tareas más complejas asignadas a los LLMs presentan menor eficiencia.

Se identificaron diferencias significativas en la eficiencia de las estrategias de diseño de prompts entre distintos LLMs. Los modelos más avanzados, como el Meta-Llama 3.1-70B-Instruct y Deep Seek, mostraron un mejor desempeño en tareas de análisis textual académico cuando se emplearon estrategias de prompt adecuadas.

Como trabajo futuro, se pretende el diseño y la evaluación de prompts en una gama más amplia de LLMs. Además, se plantea la extensión de las pruebas a una variedad de elementos constitutivos de una tesis de licenciatura, incluyendo objetivos específicos, hipótesis, justificación, antecedentes y otros componentes relevantes. Esta expansión del trabajo de investigación permitirá una comprensión más profunda de la efectividad de las estrategias de prompt en la evaluación y generación de contenido académico en diferentes contextos y niveles de complejidad.

## AGRADECIMIENTOS

Expresamos nuestro agradecimiento al asesoramiento externo: Dr. Adrián Pastor López Monroy. Investigador SNII del Centro de Investigación en Matemáticas, A.C. (CIMAT); Dr. Carlos Toxtli Hernández. Profesor Investigador de la Universidad de Clemson, USA; Dr. José Antonio Vázquez López, Investigador SNII1. Tecnológico Nacional de México; Dra. Ana Marcela Herrera Navarro, Investigadora de la Universidad Autónoma de Querétaro.

## REFERENCIAS

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. & Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems* 30 (pp. 5998-6008). (2017).
- [2] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P. & Amodei, D. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems* 33. doi.org/10.48550/arXiv.2005.14165 (pp. 1877-1901). (2020).
- [3] Sivarajkumar S, Kelley M, Samolyk-Mazzanti A, Visweswaran S, Wang Y. An Empirical Evaluation of Prompting Strategies for Large Language Models in Zero-Shot Clinical Natural Language Processing: Algorithm Development and Validation Study. *JMIR Med Inform* 2024;12:e55318 URL: <https://medinform.jmir.org/2024/1/e55318> DOI: 10.2196/55318. (2024).

- [4] Li, J., & Li, X. Towards a Deep Understanding of Language Models' Behaviors via Prompt-based Exploration. arXiv preprint arXiv:2104.08319.(2021).
- [5] Schick, T., & Schütze, H. It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners. In EMNLP. (2021).
- [6] Zhou, D., & Xu, P. On the Role of Prompting in Few-Shot Learning of Language Models. In arXiv preprint arXiv:2012.15840. (2020).
- [7] Wei, J., et al. Finetuned Language Models Are Zero-Shot Learners. In arXiv preprint arXiv:2109.01652. (2021).
- [8] Raffel, C., et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. In JMLR. (2019).
- [9] McCoy, R. T., et al. Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference. In ACL. (2021).
- [10] Bontcheva, K., & Maynard, D. Trends and Issues in Human Language Technology. In Proceedings of the 12th Language. (2019). Resources and Evaluation Conference (pp. 6785-6792).
- [11] Hovy, D., & Spruit, S. L. The Social Impact of Natural Language Processing. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (pp. 14-21). (2016).
- [12] Kaplan, J., McCann, B., & Naik, M. The Unreasonable Effectiveness of Language Models for Extracting Information from Natural Language Text. In arXiv preprint arXiv:2004.10246. (2020).
- [13] Dodge, J., Gururaj, A., & Bangalore, S. Evaluating the Impact of Context Length on Open-Domain Question Answering. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 3249-3259). (2021).

## ANEXOS

### Anexo 1

**Sección 1.** Prompt Zero Shot para identificar el Verbo en infinitivo en el Objetivo General.

Identifica el verbo principal en modo infinitivo en el siguiente objetivo general de un trabajo de tesis de nivel licenciatura. En caso de contener un verbo principal en modo infinitivo asigna calificación = 1 y si no contiene un verbo principal en modo infinitivo asigna calificación = 0. El objetivo general es: " "

**Sección 2.** Prompt Zero Shot para identificar el Tema de investigación en el Objetivo General.

Identifica en el siguiente objetivo de tesis de licenciatura el tema central de investigación y el área o disciplina académica principal. Si está bien definido el tema central de la investigación y el área o disciplina académica principal asigna una calificación=1, de lo contrario asigna una calificación=0. El objetivo a analizar es: " "

**Sección 3.** Prompt Zero Shot para identificar el Alcance o contexto de la investigación en el Objetivo General.

Encuentra en el siguiente objetivo general de una tesis de licenciatura el contexto de la investigación, ámbito o campo temático general en el que se enmarca la investigación, así como la delimitación geográfica o temporal. Si identificas el contexto, ámbito o campo temático, delimitación geográfica o temporal asigna una calificación=1, de lo contrario asigna una calificación=0. El objetivo a analizar es: " "

**Sección 4.** Prompt Zero Shot para identificar el Propósito o finalidad de la investigación en el Objetivo General.

Reconoce en el siguiente objetivo general de una tesis de licenciatura el propósito fundamental de la investigación. Si identificas el propósito fundamental de la investigación asigna una calificación =1, de lo contrario asigna una calificación=0.

El objetivo general a analizar es: " "

**Sección 5.** Prompt Zero Shot para identificar la Congruencia gramatical de la investigación en el Objetivo General.

Analiza el siguiente objetivo general de una tesis de licenciatura y asigna un puntaje de 1 si tiene congruencia gramatical, o un puntaje de 0 si no la tiene. El objetivo a analizar es: " "

## **Anexo 2**

**Sección 1.** Prompt CoT para identificar el Verbo en infinitivo en el Objetivo General.

Eres un experto investigador y extremadamente crítico. Analiza el siguiente objetivo general de un trabajo de tesis de licenciatura y utiliza el enfoque paso a paso para identificar el verbo principal en modo infinitivo. Sigue las siguientes instrucciones:

1. Identifica todos los verbos en modo infinitivo en el texto. Recuerda que los verbos en modo infinitivo son verbos con terminación "ar", "er", "ir"

2. Determina cuál es el verbo principal que representa la acción central del objetivo, recuerda que el verbo principal es la primera palabra terminada en "ar", "er", "ir" y no es un sustantivo.
3. Si hay otros verbos en infinitivo, considéralos como secundarios.
4. Considera que debe existir sólo un verbo en modo infinitivo al principio del objetivo general.
5. En caso de contener sólo un verbo principal en modo infinitivo al inicio del objetivo general asigna calificación = 1 y en caso contrario asigna calificación = 0.

El objetivo general es: " "

### **Sección 2.** Prompt CoT para identificar el Tema de investigación en el Objetivo General.

Eres un experto investigador y extremadamente crítico, analiza cuidadosamente el siguiente objetivo general de una tesis de licenciatura y utiliza un enfoque paso a paso para identificar el tema de investigación principal. Sigue estas instrucciones:

1. Reconoce el enfoque específico: Dentro del campo identificado, determina el área de especialización o el tema particular que se está abordando.
2. Determina el objeto de estudio: Identifica el fenómeno, población, sistema o elemento principal que está siendo investigado.
3. Recuerda que debe estar considerado sólo un objeto de estudio dentro del objetivo general de la investigación.
4. Identifica el complemento directo (utilizando las palabras tal como están escritas en el objetivo general analizado).

Ahora, aplica estos pasos al siguiente objetivo general y proporciona una calificación =1 si está bien definido el Complemento directo o tema central de la investigación, de lo contrario asigna una calificación=0. El objetivo a analizar es: " "

### **Sección 3.** Prompt CoT para identificar el Alcance o contexto de la investigación en el Objetivo General.

Eres un experto investigador y extremadamente crítico, analiza el siguiente objetivo general de una tesis de licenciatura e identifica los elementos clave relacionados con el contexto y el alcance de la investigación utilizando el enfoque paso a paso, siguiendo las siguientes instrucciones:

1. Identifica el complemento circunstancial de lugar es decir la delimitación geográfica o poblacional que define los límites de este estudio.



2. Identifica el complemento circunstancial de modo es decir, las manera o la forma en cómo se llegará a cumplir el objetivo.

Ahora, aplica estos pasos al siguiente objetivo general y proporciona una calificación =1 si está bien definido el Complemento circunstancial de lugar o el complemento circunstancial de modo, de lo contrario asigna una calificación=0. El objetivo a analizar es: " "

**Sección 4.** Prompt CoT para identificar el Propósito o finalidad de la investigación en el Objetivo General.

Eres un experto investigador y extremadamente crítico, analiza el siguiente objetivo general de una tesis de licenciatura para identificar su propósito o finalidad fundamental. Utiliza el enfoque paso a paso. Sigue las siguientes instrucciones

1. Identifica el complemento circunstancial de finalidad, es decir cuál es la intención fundamental de esta investigación.
2. Reconoce los resultados esperados o beneficios que se esperan.
3. Recuerda que estos aportes deberán estar textualmente definidos en el objetivo general analizado
4. Recuerda que algunas conjunciones que son usadas comúnmente para identificar el aporte pueden ser: "para", "a fin de", "con el fin de", " con la intención de".
5. Considera que el complemento circunstancial de finalidad no es el tema de la investigación.

Si identificas el propósito fundamental de la investigación o complemento circunstancial de finalidad asigna una calificación =1, de lo contrario asigna una calificación=0.

Objetivo general a analizar: " "

### **Sección 5.**

Prompt CoT para identificar la Congruencia gramatical de la investigación en el Objetivo General.

Eres un experto investigador y extremadamente crítico, analiza el siguiente objetivo general de una tesis de licenciatura y determina si tiene congruencia gramatical. Sigue estos pasos:

1. Identifica el sujeto que es quien o qué realiza la acción principal en el objetivo.
2. Identifica el verbo que es la acción principal que se propone realizar en la investigación.
3. Revisa que el verbo concuerde con el sujeto en número y persona.
4. Identifica el objeto, que es lo que se está analizando, desarrollando o evaluando.
5. Reconoce si el objeto de la acción está claramente definido.

6. Identifica el complemento circunstancial de modo.
7. Identifica el complemento circunstancial de lugar.
8. Identifica el complemento circunstancial de la finalidad.
9. Analiza si todos los elementos se encuentran correctamente integrados (sujeto, verbo, objeto, complementos circunstanciales de modo, lugar y complemento circunstancial de finalidad).
10. Revisa la estructura gramatical y observa si la oración tiene una construcción coherente y sin errores gramaticales.
11. Evalúa la congruencia:  
Si todos los elementos gramaticales están correctamente estructurados, asigna una calificación=1 Si hay incongruencias o errores gramaticales, asigna una calificación= 0.  
Objetivo general a analizar: " "

### **Anexo 3**

**Sección 1.** Prompt Few Shot-CoT para identificar el Verbo en infinitivo en el Objetivo General.

Question:

Eres un experto investigador y eres extremadamente crítico, considerando el siguiente objetivo "Optimizar métodos centrales y Generar caminos para la empresa ". Dime si está bien definido verbo principal, recuerda solo vas a analizar si está bien definido el verbo principal

Considera la siguiente cadena de pensamiento

1. Encuentra el verbo principal
2. Solo debe de existir un verbo principal
3. El verbo principal debe de estar en infinitivo, no en sustantivo
4. El verbo es medible
5. En caso de estar bien definido el verbo, calificación = 1 y si no está bien definido calificación = 0

Answer:

Analicemos el objetivo dado según la cadena de pensamiento proporcionada:

1. Encuentra el verbo principal

Puedo identificar dos verbos principales: Optimizar y Generar

2. Solo debe de existir un verbo principal

Hay dos verbos principales identificados, por lo que este criterio no se cumple

3. El verbo principal debe de estar en infinitivo, no en sustantivo

Ambos verbos principales ("Optimizar" y "Generar") están en infinitivo

4. El verbo es medible

Tanto "Optimizar" como "Generar" son verbos medibles.

5. En caso de estar bien definido el verbo, calificación = 1 y si no está bien definido calificación = 0

Dado que hay más de un verbo principal:

Calificación = 0

Question:

Eres un experto investigador y eres extremadamente crítico, considerando el siguiente objetivo "Analizar la estrecha relación entre el nivel de productividad y el nivel de satisfacción del empleado". Dime si está bien definido verbo principal, recuerda solo vas a analizar si está bien definido el verbo principal

Considera la siguiente cadena de pensamiento

1. Encuentra el verbo principal
2. Solo debe de existir un verbo principal
3. El verbo principal debe de estar en infinitivo, no en sustantivo
4. El verbo es medible
5. En caso de estar bien definido el verbo, calificación = 1 y si no está bien definido calificación = 0

Answer:

Analicemos el objetivo dado según la cadena de pensamiento proporcionada:

1. Encuentra el verbo principal

Puedo identificar el siguiente verbo principal: Analizar

2. Solo debe de existir un verbo principal

Hay solo un verbo principal, por lo que este criterio se cumple

3. El verbo principal debe de estar en infinitivo, no en sustantivo

El verbo "analizar" está en infinitivo, por lo que este criterio se cumple

4. El verbo es medible

El verbo "analizar" es medible, por lo que este criterio se cumple.

5. En caso de estar bien definido el verbo, calificación = 1 y si no está bien definido calificación = 0

Dado que el verbo principal cumple con todos los criterios:

Calificación = 1

Question:

Eres un experto investigador y eres extremadamente crítico, considerando el siguiente objetivo "Posicionar y penetrar en el mercado al equipo de futbol Titanes de Querétaro para ser reconocido y lograr tener presencia en el municipio haciendo que los jóvenes se interesen en formar parte del mismo.". Dime si está bien definido verbo principal, recuerda solo vas a analizar si está bien definido el verbo principal

Considera la siguiente cadena de pensamiento

1. Encuentra el verbo principal
2. Solo debe de existir un verbo principal
3. El verbo principal debe de estar en infinitivo, no en sustantivo
4. El verbo es medible
5. En caso de estar bien definido el verbo, calificación = 1 y si no está bien definido calificación = 0

Answer:

Analicemos el objetivo dado según la cadena de pensamiento proporcionada:

1. Encuentra el verbo principal

Puedo identificar dos verbos principales: Posicionar y Penetrar

2. Solo debe de existir un verbo principal

Hay dos verbos principales identificados, por lo que este criterio no se cumple

3. El verbo principal debe de estar en infinitivo, no en sustantivo

Ambos verbos principales ("Posicionar" y "Penetrar") están en infinitivo

4. El verbo es medible

Tanto "Posicionar" como "Penetrar" son verbos medibles.

5. En caso de estar bien definido el verbo, calificación = 1 y si no está bien definido calificación = 0

Dado que hay más de un verbo principal:

Calificación = 0

Question:

Eres un experto investigador y eres extremadamente crítico, considerando el siguiente objetivo "Programa de mantenimiento para los equipos de control numérico". Dime si está bien definido verbo principal, recuerda solo vas a analizar si está bien definido el verbo principal

Considera la siguiente cadena de pensamiento

1. Encuentra el verbo principal

2. Solo debe de existir un verbo principal

3. El verbo principal debe de estar en infinitivo, no en sustantivo

4. El verbo es medible

5. En caso de estar bien definido el verbo, calificación = 1 y si no está bien definido calificación = 0

Answer:

Analicemos el objetivo dado según la cadena de pensamiento proporcionada:

1. Encuentra el verbo principal

Puedo identificar el siguiente verbo principal: mantenimiento

2. Solo debe de existir un verbo principal

En este caso, solo hay un verbo principal, "mantenimiento"

3. El verbo principal debe de estar en infinitivo, no en sustantivo

El verbo "mantenimiento" no está en infinitivo, es un sustantivo; la forma correcta sería "mantener".

4. El verbo es medible

"mantener" puede ser medible, pero como está actualmente en sustantivo ("mantenimiento"), no cumple con este criterio de manera directa

5. En caso de estar bien definido el verbo, calificación = 1 y si no está bien definido calificación = 0

El verbo principal no está bien definido porque no está en infinitivo.

Calificación = 0

Question:

Eres un experto investigador y eres extremadamente crítico, considerando el siguiente objetivo "Implementar la metodología 5E en el área de fábrica de compuestos con el fin de mejorar la productividad de la empresa y la calidad de sus productos.". Dime si está bien definido verbo principal, recuerda solo vas a analizar si está bien definido el verbo principal

Considera la siguiente cadena de pensamiento

1. Encuentra el verbo principal

2. Solo debe de existir un verbo principal

3. El verbo principal debe de estar en infinitivo, no en sustantivo

4. El verbo es medible

5. En caso de estar bien definido el verbo, calificación = 1 y si no está bien definido calificación = 0

**Sección 2.** Prompt Few Shot-CoT para identificar el Tema de investigación en el Objetivo General.

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Identificar las acciones de mejora Evaluar la situación actual de la planta en cuanto a la generación de material no conforme proveniente de todos los procesos de Línea #2.". Dime si está bien definido el tema de investigación, recuerda solo vas a evaluar si está bien definido el tema de investigación

Considera la siguiente cadena de pensamiento

1. Encuentra el Verbo de acción

2. Encuentra el objeto de estudio

3. Revisa que ni haya más de un objeto de estudio

4. En caso de estar bien definido el tema de investigación, calificación = 1 y si no está bien definido calificación = 0

Answer:

Voy a evaluar si el tema de investigación está bien definido siguiendo la cadena de pensamiento propuesta.

Encuentra el Verbo de acción:

El verbo de acción principal es "Identificar".

Encuentra el objeto de estudio:

El objeto de estudio es "las acciones de mejora".

Revisa que no haya más de un objeto de estudio:

Aquí surge un problema. Aunque el verbo principal es "Identificar" y el objeto principal es "las acciones de mejora", la frase incluye otro objetivo: "Evaluar la situación actual de la planta". Esto introduce un segundo objeto de estudio, lo cual puede generar confusión sobre el enfoque principal de la investigación.

Calificación:

Debido a la presencia de dos objetos de estudio distintos ("acciones de mejora" y "situación actual de la planta"), el tema de investigación no está bien definido.

Calificación = 0

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Analizar el crédito hipotecario INFONAVIT para determinar si es o no una adecuada opción para que la población de San Luis de la Paz lo adquiera". Dime si está bien definido el tema de investigación, recuerda solo vas a evaluar si está bien definido el tema de investigación Considera la siguiente cadena de pensamiento

1. Encuentra el Verbo de acción
2. Encuentra el objeto de estudio
3. Revisa que ni haya más de un objeto de estudio
4. En caso de estar bien definido el tema de investigación, calificación = 1 y si no está bien definido calificación = 0

Answer:

Voy a evaluar si el tema de investigación está bien definido siguiendo la cadena de pensamiento propuesta.

Encuentra el Verbo de acción:

El verbo de acción principal es "Analizar".

Encuentra el objeto de estudio:

El objeto de estudio es "el crédito hipotecario INFONAVIT".

Revisa que no haya más de un objeto de estudio:

Efectivamente, solo hay un objeto de estudio principal, que es el crédito hipotecario INFONAVIT.

Calificación:

El tema está bien definido, ya que tiene un verbo de acción claro y un objeto de estudio único. Además, especifica el propósito del análisis (determinar si es una adecuada opción) y el contexto geográfico (para la población de San Luis de la Paz).

Calificación = 1

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Elaboración precisa de piezas y optimización del rendimiento de la prensa hidráulica: Garantizar la fabricación precisa de las piezas necesarias para mantener un funcionamiento óptimo de la prensa hidráulica. Esto implica asegurar que las piezas cumplan con los estándares de calidad y tolerancia requeridos para su correcto funcionamiento.". Dime si está bien definido el tema de investigación, recuerda solo vas a evaluar si esta bien definido el tema de investigación

Considera la siguiente cadena de pensamiento

1. Encuentra el Verbo de acción
2. Encuentra el objeto de estudio
3. Revisa que ni haya más de un objeto de estudio
4. En caso de estar bien definido el tema de investigación, calificación = 1 y si no está bien definido calificación = 0

Answer:

Voy a evaluar si el tema de investigación está bien definido siguiendo la cadena de pensamiento propuesta.

Encuentra el Verbo de acción:

El verbo principal es "Elaboración" y "Optimización", que son adecuados para una investigación.

Encuentra el objeto de estudio:

El objeto de estudio es la "prensa hidráulica" y específicamente la "fabricación precisa de las piezas necesarias" para su funcionamiento óptimo.

Revisa que no haya más de un objeto de estudio:

Hay dos objetos de estudio relacionados pero distintos:

\* La "elaboración precisa de piezas"

\* La "optimización del rendimiento de la prensa hidráulica"

Aunque están relacionados, son dos aspectos diferentes que podrían ser abordados de manera independiente.

Calificación:

El tema de investigación no está bien definido debido a la presencia de dos objetos de estudio relacionados pero distintos. Sería recomendable redefinir

el objetivo para centrarse en uno de los aspectos o clarificar cómo se relacionan entre sí los dos objetos de estudio.

Calificación = 0

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Analizar estrategias de mejora continua para reducir el scrap dentro del proceso productivo de cutting ("Rebabeo"), en piezas de caucho del área de RUBBER ("Caucho"), en DONGA HWASUNG MEXICO S.A. DE C.V". Dime si está bien definido el tema de investigación, recuerda solo vas a evaluar si está bien definido el tema de investigación

Considera la siguiente cadena de pensamiento

1. Encuentra el Verbo de acción
2. Encuentra el objeto de estudio
3. Revisa que ni haya más de un objeto de estudio
4. En caso de estar bien definido el tema de investigación, calificación = 1 y si no está bien definido calificación = 0

**Sección 3.** Prompt Few Shot-CoT para identificar el Alcance o contexto de la investigación en el Objetivo General

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Determinar el tiempo promedio para el cambio de formato 60x60, en la prensa hidráulica 3800.". Dime si está bien definido el contexto de investigación, recuerda solo vas a evaluar si está bien definido el contexto de investigación Considera la siguiente cadena de pensamiento

1. Encentra el contexto donde se realizará la investigación
2. Identificación de elementos clave del contexto
3. analiza de la especificidad del contexto
4. Determina si el contexto está relacionado con el objetivo
5. En caso de estar bien definido el contexto, calificación = 1 y si no está bien definido calificación = 0

Answer:

1. Encentra el contexto donde se realizará la investigación  
Prensa hidráulica específica (modelo 3800)
2. Identificación de elementos clave del contexto  
Prensa hidráulica modelo 3800, Cambio de formato 60x60, Medición de tiempo
3. Analiza de la especificidad del contexto



El contexto es bastante específico, mencionando un modelo concreto de máquina (prensa hidráulica 3800)

4. Determina si el contexto está relacionado con el objetivo

El contexto está directamente relacionado con el objetivo

5. En caso de estar bien definido el contexto, calificación = 1 y si no está bien definido calificación = 0

El contexto de la investigación está bien definido porque se especifica claramente la máquina en la que se realizará la investigación (prensa hidráulica 3800).

Calificación =1

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Optimizar métodos centrales y Generar caminos para la empresa.". Dime si está bien definido el contexto de investigación, recuerda solo vas a evaluar si está bien definido el contexto de investigación Considera la siguiente cadena de pensamiento

1. Encentra el contexto donde se realizará la investigación

2. Identificación de elementos clave del contexto

3. analiza de la especificidad del contexto

4. Determina si el contexto está relacionado con el objetivo

5. En caso de estar bien definido el contexto, calificación = 1 y si no está bien definido calificación = 0

Answer:

1. Encentra el contexto donde se realizará la investigación

No se proporciona un contexto específico para la investigación. Falta información sobre el tipo de empresa, sector, tamaño, ubicación geográfica, etc.

2. Identificación de elementos clave del contexto

Debido a la falta de contexto, no es posible identificar elementos clave.

3. Analiza de la especificidad del contexto

El contexto es inexistente, por lo tanto, no hay especificidad que analizar.

4. Determina si el contexto está relacionado con el objetivo

Sin un contexto definido, no se puede establecer una relación clara con el objetivo propuesto.

5. En caso de estar bien definido el contexto, calificación = 1 y si no está bien definido calificación = 0

Dado que no se proporciona un contexto para la investigación, el contexto no está bien definido.

Calificación =0

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Posicionar y penetrar en el mercado al equipo de futbol Titanes de Querétaro para ser reconocido y lograr tener presencia en el municipio haciendo que los jóvenes se interesen en formar parte del mismo". Dime si está bien definido el contexto de investigación, recuerda solo vas a evaluar si está bien definido el contexto de investigación Considera la siguiente cadena de pensamiento

1. Encuentra el contexto donde se realizará la investigación
2. Identificación de elementos clave del contexto
3. analiza de la especificidad del contexto
4. Determina si el contexto está relacionado con el objetivo
5. En caso de estar bien definido el contexto, calificación = 1 y si no está bien definido calificación = 0

Answer:

1. Encuentra el contexto donde se realizará la investigación  
El contexto parece ser el municipio de Querétaro, México, centrándose en el mercado deportivo local, específicamente en el fútbol juvenil.
2. Identificación de elementos clave del contexto  
Equipo de fútbol: Titanes de Querétaro, Ubicación: Municipio de Querétaro

3. Analiza de la especificidad del contexto  
El contexto es bastante específico en cuanto a la ubicación (Querétaro) y el público objetivo (jóvenes)
4. Determina si el contexto está relacionado con el objetivo  
El contexto está directamente relacionado con el objetivo de posicionar y penetrar en el mercado al equipo Titanes de Querétaro.
5. En caso de estar bien definido el contexto, calificación = 1 y si no está bien definido calificación = 0

El contexto está relacionado con el objetivo

Calificación =1

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Posicionar y penetrar en el mercado al equipo de futbol Titanes de Querétaro para ser reconocido y lograr tener presencia en el municipio haciendo que los jóvenes se interesen en formar parte del mismo". Dime si está bien definido

el contexto de investigación, recuerda solo vas a evaluar si está bien definido el contexto de investigación Considera la siguiente cadena de pensamiento

1. Encuentra el contexto donde se realizará la investigación
2. Identificación de elementos clave del contexto
3. analiza de la especificidad del contexto
4. Determina si el contexto está relacionado con el objetivo
5. En caso de estar bien definido el contexto, calificación = 1 y si no está bien definido calificación = 0

**Sección 4.** Prompt Few Shot-CoT para identificar el Propósito o finalidad de la investigación en el Objetivo General.

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Optimizar métodos centrales y Generar caminos para la empresa ". Dime si se identifica el propósito del siguiente objetivo, recuerda solo vas a evaluar se identifica el propósito del objetivo, genera tu respuesta considerando la siguiente cadena de pensamiento

1. Identifica el verbo o acción principal
2. Determina el objeto de estudio
3. Identifica el contexto o ámbito
4. Identifica el resultado esperado o contribución
5. Sintetiza este propósito
6. En caso de identificar claramente el propósito calificación = 1 y si no se identifica calificación = 0

Answer:

1. Identifica el verbo o acción principal  
Hay dos verbos principales: "Optimizar" y "Generar"

2. Determina el objeto de estudio  
métodos centrales", "caminos"

3. Identifica el contexto o ámbito  
"para la empresa"

4. Identifica el resultado esperado o contribución

No se especifica claramente un resultado esperado o contribución concreta.

5. Sintetiza este propósito

El propósito parece ser mejorar procesos internos clave y crear nuevas estrategias o direcciones para la empresa, pero carece de especificidad en cuanto a los resultados esperados o el impacto deseado.

6. En caso de identificar claramente el propósito calificación = 1 y si no se identifica calificación = 0

No se identifica claramente el propósito del objetivo. La falta de especificidad y contexto hace que el objetivo sea confuso y poco útil en su forma actual.

Calificación = 0

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Analizar la estrecha relación entre el nivel de productividad y el nivel de satisfacción del empleado". Dime si se identifica el propósito del siguiente objetivo, recuerda solo vas a evaluar se identifica el propósito del objetivo, genera tu respuesta considerando la siguiente cadena de pensamiento

la siguiente cadena de pensamiento

1. Identifica el verbo o acción principal
2. Determina el objeto de estudio
3. Identifica el contexto o ámbito
4. Identifica el resultado esperado o contribución
5. Sintetiza esté propósito
6. En caso de identificar claramente el propósito calificación = 1 y si no se identifica calificación = 0

Answer

1. Identifica el verbo o acción principal

El verbo principal es "Analizar".

2. Determina el objeto de estudio

El objeto de estudio es "la estrecha relación entre el nivel de productividad y el nivel de satisfacción del empleado".

3. Identifica el contexto o ámbito

El contexto parece ser el ámbito laboral o empresarial, aunque no se especifica explícitamente un sector o tipo de organización en particular.

4. Identifica el resultado esperado o contribución

No se menciona explícitamente un resultado esperado o una contribución específica. El objetivo no indica qué se hará con el análisis o cómo se aplicará.

5. Sintetiza esté propósito

El propósito es examinar cómo se relacionan estrechamente la productividad y la satisfacción de los empleados en un entorno laboral,

6. En caso de identificar claramente el propósito calificación = 1 y si no se identifica calificación = 0

Falta especificar el resultado esperado o la contribución de este análisis

Calificación = 0

**Question:**

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Fortalecer el proceso de ensamble mediante la estrategia de capacitación de ayudas visuales para brindar al personal de producción en el área de ensamble, una mejor información y entendimiento sobre las actividades que influyen en su trabajo.". Dime si se identifica el propósito del siguiente objetivo, recuerda solo vas a evaluar se identifica el propósito del objetivo, genera tu respuesta considerando la siguiente cadena de pensamiento

la siguiente cadena de pensamiento

1. Identifica el verbo o acción principal
2. Determina el objeto de estudio
3. Identifica el contexto o ámbito
4. Identifica el resultado esperado o contribución
5. Sintetiza este propósito
6. En caso de identificar claramente el propósito calificación = 1 y si no se identifica calificación = 0

**Answer**

1. Identifica el verbo o acción principal

El verbo principal es "Fortalecer".

2. Determina el objeto de estudio

El objeto de estudio es "el proceso de ensamble".

3. Identifica el contexto o ámbito

El contexto es "el área de ensamble" dentro de un entorno de producción.

4. Identifica el resultado esperado o contribución

El resultado esperado es "brindar al personal de producción una mejor información y entendimiento sobre las actividades que influyen en su trabajo".

5. Sintetiza este propósito

El propósito es mejorar el proceso de ensamble mediante la capacitación del personal de producción, utilizando ayudas visuales para aumentar su comprensión de las actividades laborales.

6. En caso de identificar claramente el propósito calificación = 1 y si no se identifica calificación = 0

El propósito del objetivo se identifica claramente. Se establece una acción concreta (fortalecer), un objeto de estudio específico (proceso de ensamble), un contexto definido (área de ensamble=

Calificación = 1

**Question:**

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Analizar la estrecha relación entre el nivel de productividad y el nivel de satisfacción del empleado". Dime si se identifica el propósito del siguiente objetivo, recuerda solo vas a evaluar se identifica el propósito del objetivo, genera tu respuesta considerando la siguiente cadena de pensamiento

la siguiente cadena de pensamiento

1. Identifica el verbo o acción principal
2. Determina el objeto de estudio
3. Identifica el contexto o ámbito
4. Identifica el resultado esperado o contribución
5. Sintetiza este propósito
6. En caso de identificar claramente el propósito calificación = 1 y si no se identifica calificación = 0

**Sección 5.** Prompt Few Shot-CoT para identificar la Congruencia gramatical de la investigación en el Objetivo General.

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Determinar el tiempo promedio para el cambio de formato 60x60, en la prensa hidráulica 3800.". Dime si existe congruencia gramatical en el siguiente objetivo, recuerda solo vas a evaluar si existe congruencia gramatical en la investigación, genera tu respuesta considerando la siguiente cadena de pensamiento

1. Examinemos la estructura general de la oración
2. Identifiquemos las partes principales
3. Analicemos la congruencia entre estas partes
4. Evaluemos la puntuación
5. En caso de existir congruencia gramatical, calificación = 1 y si no está bien definido calificación = 0

Answer:

1. Examinemos la estructura general de la oración  
La oración es una frase infinitiva que expresa un objetivo. Tiene una estructura simple y directa.
2. Identifiquemos las partes principales

Verbo principal: "Determinar", Objeto directo: "el tiempo promedio",  
Complemento circunstancial de finalidad: "para el cambio de formato 60x60"  
Complemento circunstancial de lugar: "en la prensa hidráulica 3800"

3. Analicemos la congruencia entre estas partes

Las partes de la oración están correctamente relacionadas entre sí.

4. Evaluemos la puntuación

La puntuación es correcta

5. En caso de existir congruencia gramatical, calificación = 1 y si no está bien definido calificación = 0

Dado que existe congruencia gramatical en la oración y está bien definida calificación = 1.

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Optimizar métodos centrales y Generar caminos para la empresa.". Dime si existe congruencia gramatical en el siguiente objetivo, recuerda solo vas a evaluar si existe congruencia gramatical en la investigación, genera tu respuesta considerando la siguiente cadena de pensamiento

1. Examinemos la estructura general de la oración

2. Identifiquemos las partes principales

3. Analicemos la congruencia entre estas partes

4. Evaluemos la puntuación

5. En caso de existir congruencia gramatical, calificación = 1 y si no está bien definido calificación = 0

Answer:

1. Examinemos la estructura general de la oración

El objetivo presentado es "Optimizar métodos centrales y Generar caminos para la empresa." La estructura general consiste en dos acciones separadas por la conjunción "y".

2. Identifiquemos las partes principales

La oración tiene dos partes principales: "Optimizar métodos centrales" y "Generar caminos para la empresa". Ambas son oraciones coordinadas por la conjunción "y", cada una con su propio verbo y complemento.

3. Analicemos la congruencia entre estas partes

Las dos partes de la oración están gramaticalmente correctas de manera individual. Sin embargo, cuando se unen mediante la conjunción "y", no existe una clara conexión lógica o semántica entre "Optimizar métodos centrales" y

"Generar caminos para la empresa". Esto genera cierta incongruencia en cuanto al sentido global del objetivo.

4. Evaluemos la puntuación

La oración carece de puntuación interna, lo cual es adecuado dado que está compuesta por dos cláusulas coordinadas. No hay necesidad de añadir comas u otros signos de puntuación adicionales.

5. En caso de existir congruencia gramatical, calificación = 1 y si no está bien definido calificación = 0

Aunque la oración es gramaticalmente correcta en cuanto a estructura y puntuación, la falta de una conexión clara y lógica entre las dos acciones genera incongruencia semántica.

calificación = 0.

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Implementar la metodología 5E en el área de fábrica de compuestos con el fin de mejorar la productividad de la empresa y la calidad de sus productos". Dime si existe congruencia gramatical en el siguiente objetivo, recuerda solo vas a evaluar si existe congruencia gramatical en la investigación, genera tu respuesta considerando la siguiente cadena de pensamiento

1. Examinemos la estructura general de la oración
2. Identifiquemos las partes principales
3. Analicemos la congruencia entre estas partes
4. Evaluemos la puntuación
5. En caso de existir congruencia gramatical, calificación = 1 y si no está bien definido calificación = 0

Answer:

1. Examinemos la estructura general de la oración  
El objetivo es "Implementar la metodología 5E en el área de fábrica de compuestos con el fin de mejorar la productividad de la empresa y la calidad de sus productos." La oración es compleja, pero sigue una estructura lineal que se entiende claramente.
2. Identifiquemos las partes principales  
La oración contiene un verbo principal ("Implementar"), un complemento directo ("la metodología 5E en el área de fábrica de compuestos"), y un



propósito final ("con el fin de mejorar la productividad de la empresa y la calidad de sus productos").

3. Analicemos la congruencia entre estas partes

Las partes están bien relacionadas entre sí. El verbo "Implementar" es adecuado y el complemento directo especifica claramente qué se va a implementar y dónde. El propósito final se conecta lógicamente con la acción principal, añadiendo una razón clara para la implementación de la metodología 5E.

4. Evaluemos la puntuación

La puntuación es correcta. No hay comas innecesarias y la estructura de la oración fluye de manera natural. La frase "con el fin de" introduce adecuadamente la cláusula final, sin necesidad de ajustes.

5. En caso de existir congruencia gramatical, calificación = 1 y si no está bien definido calificación = 0

La oración presenta congruencia gramatical en todos sus elementos.

Calificación = 1.

Question:

Eres un experto investigador eres extremadamente crítico y vas a estructurar tu opinión en forma de cadena de pensamiento del siguiente objetivo "Programa de mantenimiento para los equipos de control numérico". Dime si existe congruencia gramatical en el siguiente objetivo, recuerda solo vas a evaluar si existe congruencia gramatical en la investigación, genera tu respuesta considerando la siguiente cadena de pensamiento

1. Examinemos la estructura general de la oración

2. Identifiquemos las partes principales

3. Analicemos la congruencia entre estas partes

4. Evaluemos la puntuación

5. En caso de existir congruencia gramatical, calificación = 1 y si no está bien definido calificación = 0

## Anexo 4

### Evaluación Global del Objetivo General de la investigación Prompt Árbol de Pensamiento

Tienes cinco expertos investigadores que son muy críticos están evaluando el siguiente objetivo de tesis "Presentar la mejor estrategia de inteligencia emocional para conocer los procesos de mejora continua en las relaciones interpersonales en la tienda de conveniencia Oxxo Vicente Guerrero"

El primero va a revisar que el verbo esté bien definido, si está bien definido el verbo colocará un punto de evaluación, considerará los siguientes puntos para dar una respuesta

1. Solo hay un verbo principal
2. El Verbo principal debe de estar en infinitivo, no en sustantivo
3. El Verbo principal esta al principio de la Oración

El segundo va a revisar que el tema esté bien definido, si está bien definido el tema colocará un punto de evaluación, considerará los siguientes puntos para dar una respuesta

1. Existe un objeto de estudio
2. No hay más de un objeto de estudio

El tercero va a revisar que el contexto esté bien definido, si está bien definido el contexto colocará un punto de evaluación, considerará los siguientes puntos para dar una respuesta

1. Existe un contexto donde se realizará la investigación
2. Identificará los elementos clave del contexto
3. Analizará la especificidad del contexto y determinará si el contexto está relacionado con el objetivo

El cuarto va a evaluar que el propósito esté bien definido si está bien definido el propósito colocará un punto de evaluación, considerará los siguientes puntos para dar una respuesta

1. Existe el objeto de estudio
2. Existe un resultado esperado o contribución

El quinto va a evaluar que la congruencia gramatical definido considerando la siguiente cadena de pensamiento, considerará los siguientes puntos para dar una respuesta

1. Examinará la estructura general de la oración
2. Identificarás las partes principales de la oración
3. Analizará la congruencia entre estas partes
4. Evaluará la puntuación

Al final los investigadores discutirán sobre la calidad de las evaluaciones de los otros investigadores, llegaran a una conclusión de cuantos puntos merece cada apartado, sumaran los puntos para entregar una evaluación final

Si la Evaluación superior o igual a 4 Se colocará Calificación = 1 en caso contrario, Calificación = 0

Como respuesta solo debes de generar resumen de 50 palabras donde incluyas la calificación, no intentes hacerlo en forma de código de programación.



# CONSTRUCCIÓN DE MODELOS DE LENGUAJE TIPO BERT PARA EL ESPAÑOL MEXICANO

**Luis G. Ruiz Velázquez, Eric S. Téllez Avila, Mario Graff Guerrero**

*Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación (INFOTEC)*

*Cto. Tecnopolo Sur 112*

*CP 20326, Aguascalientes, Aguascalientes*

*Email: {luis.ruiz,eric.tellez,mario.graff}@infotec.mx*

**Daniela Moctezuma**

*Centro de Investigación en Ciencias de Información Geoespacial (CENTROGEO)*

*Cto Tecnopolo Norte 107, colonia Pocitos II,*

*CP 20313, Aguascalientes, Aguascalientes*

*Email: dmoctezuma@centrogeo.edu.mx*

**Resumen.** En este artículo se presenta un nuevo modelo de lenguaje que fue específicamente diseñado y entrenado para reconocer información regional del español de México y pueda utilizarse en tareas de Procesamiento de Lenguaje Natural. Se entrenaron dos modelos con 56 y 132 millones de parámetros respectivamente usando más de 140 millones de mensajes de la red social Twitter (ahora X). La información regional aprendida se puede corroborar en la evaluación, donde los modelos propuestos superaron a BETO al clasificar los mensajes en el estado de la república donde fueron escritos. Los modelos son de código abierto y están disponibles en la plataforma Hugging Face.

**Palabras clave:** Modelos de lenguaje, variantes mexicanas del español, regionalización de modelos

## 1 INTRODUCCIÓN

Una de las principales características que nos diferencia de las demás especies es la complejidad del lenguaje. Éste nos permitió comunicar ideas sofisticadas que fomentaron el establecimiento de las primeras sociedades. El español es el segundo idioma más hablado en el mundo, sólo después del chino, y se habla en 20 países como resultado de la colonización española iniciada en el S. XV.

Esta diversidad tanto en el territorio como en el número de hablantes, hace al español un idioma con muchos dialectos o variantes que comparten palabras

y estructuras gramaticales, pero que también evolucionaron para adaptarse a las necesidades locales.

El estudio de los lenguajes es el objetivo de la rama de la computación llamada Procesamiento del Lenguaje Natural (PLN). Las redes neuronales artificiales han sido empleadas en PLN con gran éxito creando modelos para capturar la semántica de las palabras, llamados encajes de palabras o *word embeddings*, que pueden utilizarse en tareas como traducción automática, creación de resúmenes, sistemas de preguntas y respuestas, entre muchas otras. Ejemplos de estos son presentados por Mikolov et al. [1] y Bojanowski et al. 2017 [2]. Estas técnicas se centran en representar los significados de las palabras; en Vaswani et al. 2017 [3] se presenta la arquitectura profunda llamada *Transformer* que se centra en tareas de PLN usando representaciones sensibles del contexto. Devlin et al. [4] introduce el modelo BERT (Bidirectional Encoder Representations from Transformers) que revolucionó el área, y en particular, las tareas, que toman ventaja de la semántica aprendida por el modelo, como es la clasificación de texto. Estos modelos son de gran utilidad ya que pueden ser pre-entrenados y especializados (afinados) para tareas específicas, amortizando su altísimo costo de construcción con costos relativamente bajos de especialización. El alto costo de construcción viene dado por la gran cantidad de datos que requieren para aprender, así como de la altísima complejidad del modelo.

Para el idioma inglés existe una gran cantidad de textos que pueden ser usados como entrenamiento y un gran interés por usarlos para múltiples variantes [5]. En el caso del español, aunque existen modelos como BETO [6] que pueden ser utilizados para resolver tareas de PLN, sin embargo, como se describe en Tellez et al. [7], dada la diversidad del español, es necesario tener en cuenta las diferentes variantes y modismos de la región. En el caso específico de México, los censos poblacionales con más de 129 millones de habitantes en el territorio,<sup>1</sup> poco menos de 13 millones en Estados Unidos de primera generación y más de 26 millones de segunda generación,<sup>2</sup> el español mexicano merece una especial atención para la creación de grandes modelos de lenguaje que permitan la creación de aplicaciones que sean capaces de entender de manera diferenciada las variantes de la lengua española en México, con el fin

---

<sup>1</sup> [https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2023/EAP\\_DMPO23.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2023/EAP_DMPO23.pdf)

<sup>2</sup> [https://es.wikipedia.org/wiki/Inmigraci3n\\_mexicana\\_en\\_los\\_Estados\\_Unidos](https://es.wikipedia.org/wiki/Inmigraci3n_mexicana_en_los_Estados_Unidos)

de mejorar los desempeños de los modelos en tareas que beneficien a la población en cuestión.

Nuestra contribución es la creación de modelos de lenguaje pre-entrenados basados en BERT para diferentes variantes del español Mexicano utilizando datos de Twitter para el aprendizaje del modelo. En particular, los modelos generados contienen información regional explícita empotrada que permite informar al modelo si se desea utilizar información regional, sesgar el modelo hacia alguna región, y en último caso, la de predecir de manera natural la región más probable para mensajes dados.

## 2. SOBRE EL CONJUNTO DE ENTRENAMIENTO

### 2.1 Descripción

El conjunto de entrenamiento consistió en mensajes de la red social Twitter (ahora X) obtenidos mediante su API en fechas desde diciembre de 2015 hasta febrero de 2023. Los mensajes contienen una serie de características como la fecha, el nombre del usuario, mensaje, coordenadas geoespaciales, entre muchos otros.

Se seleccionaron los mensajes escritos desde México y mediante las coordenadas geográficas, se les asignó el estado de procedencia como una etiqueta extra. En total se obtuvieron 147,171,115 mensajes de los cuales se usaron 145,827,470 para el entrenamiento y el resto para la evaluación.

### 2.2 Preprocesamiento

Antes de entrenar con los mensajes de texto, se les dió un tratamiento previo para estandarizar y eliminar algunos datos que pudieran ser sensibles. En concreto se procedió de la siguiente manera:

- Se cambiaron las menciones de usuario por `_USR`.
- Se cambiaron las direcciones de internet por `_URL`.
- Se estandarizaron expresiones como `jajaja` o `jajajaa` a sólo una repetición: `jaja`.
- Se reemplazaron `&amp;`, `&gt;` y `&lt;` por `&`, `>` y `<` respectivamente.

Las mayúsculas no se cambiaron.

Para introducir información regional, se optó por usar la siguiente convención: `<estado> _GEO <mensaje>`. Es decir, se incluyó una etiqueta con el estado de la república desde donde se publicó el mensaje, el token separador `_GEO` y luego el mensaje de texto. Esta técnica nos asegura que el modelo aprenda información relacionada con cada región, es decir, le damos una posición que se puede usar para aprender un sesgo para cada región.

La Tabla 1 muestra algunos ejemplos de mensajes que se usaron durante el entrenamiento. Se puede ver que se dejaron los *hashtags*, emoticones y las palabras mal escritas.

*Tabla 1: Ejemplos de mensajes de entrenamiento.*

Coahuila\_GEO Cómo estás amiga, nos conocemos? Soy soltero busco soltera.  
#PiedrasNegras #nava #allende #zaragoza

Tamaulipas\_GEO Ando de buenas que ya les devolví sus unfollows y métanselos por el culo ☺.

BCS\_GEO Ésa canción que cantas en silencio y la otra persona tmb. Bn raro.

Tamaulipas\_GEO Hoy es la primera vez que manejo en estado de ebriedad  
😓😓

## 2.3 Creación del vocabulario

La creación del vocabulario es una parte importante del diseño de un modelo de lenguaje. Para nuestro caso se eligió un tokenizador WordPiece<sup>3</sup> el cual fue diseñado específicamente para BERT [4]; se usó un vocabulario de 30K términos que los aprendió del conjunto de entrenamiento.

Es importante mencionar que las etiquetas que indican la entidad federativa o región se eligieron para que sean representadas con un sólo token. Esto con la finalidad de que sea más sencillo su enmascaramiento a la hora de entrenar y predecir. De esta manera, la tarea de predecir el estado de procedencia se podrá realizar reemplazando el primero token por el token de máscara *<mask>*.

## 3. ENTRENAMIENTO

### 3.1 Configuración del modelo

La arquitectura elegida para el entrenamiento fue RoBERTa [8] que está

---

<sup>3</sup>Documentación de WordPiece <https://huggingface.co/learn/nlp-course/en/chapter6/6>



disponible mediante *Hugging Face*.<sup>4</sup> Se tomaron dos configuraciones para los modelos *small*<sup>5</sup> y *large*<sup>6</sup> que se pueden consultar en la Tabla 2.

Tabla 2: Configuraciones de los modelos entrenados.

Modelo	Tamaño del encaje	Cabezas de atención	Capas ocultas	Parámetros
MX Small	512	8	6	56.5 M
MX Large	768	8	12	132.1 M

Los modelos fueron entrenados usando pesos iniciales aleatorios. El entrenamiento se realizó en dos tarjetas Nvidia TITAN RTX con 24GB de RAM; durante los meses de marzo a agosto de 2024.

### 3.2 Parámetros del entrenamiento

El entrenamiento se hizo durante un época (epoch) sobre el conjunto de datos usando el enmascaramiento de tokens. A cada mensaje se le aplicó un enmascaramiento aleatorio con 15% de probabilidad sobre cada token. La tarea consiste en que el modelo debe predecir los tokens enmascarados tomando en cuenta el resto del mensaje. Se usó una tasa de aprendizaje de  $2 \times 10^{-5}$ .

## 4. EVALUACIÓN

### 4.1 Similitud entre regiones

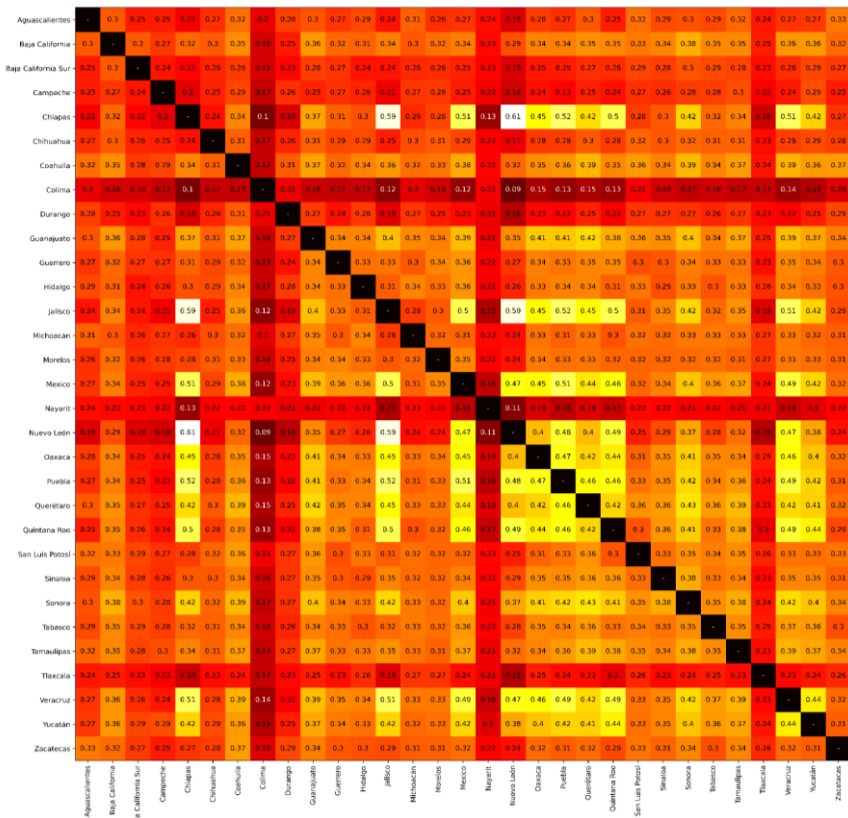
Una de las preguntas principales que queremos responder es ¿qué tanto varía la forma de escribir entre los diferentes estados del país? Los modelos entrenados nos permiten hacer esta comparación mediante las capas de atención presentes en la arquitectura de transformer.

<sup>4</sup>Sitio de Hugging Face <https://huggingface.co/>

<sup>5</sup>Disponible en: [https://huggingface.co/quillermoruz/mex\\_state](https://huggingface.co/quillermoruz/mex_state)

<sup>6</sup>Disponible en: [https://huggingface.co/quillermoruz/mex\\_large](https://huggingface.co/quillermoruz/mex_large)

Usando el modelo MX small, tomamos mensajes del conjunto de prueba para predecir su región y guardamos los tokens del mensaje con mayor influencia durante la predicción del token de estado. Para esto, usamos las cabezas de atención en la última capa de *encoder*. Con estos vocabularios, aplicamos la similitud de Dice para compararlos. La Figura 1 muestra el mapa de calor con la similitud de cada estado usando la capa de atención número 8. De la figura, podemos observar que los estados Chiapas, Jalisco, México, Nuevo León, Oaxaca, Puebla, Querétaro, Quintana Roo, Veracruz y Yucatán son todos similares entre sí. Sonora se parece a Baja California, Coahuila, Guanajuato, Sinaloa y Tamaulipas. Otros estados con vocabulario parecido son Michoacán e Hidalgo; Morelos con Guerrero; Tabasco con San Luis Potosí. Una observación interesante es que los estados de Nayarit, Colima y Tlaxcala son los menos parecidos a los demás.



*Figura 1 – Similitud entre los vocabularios de los estados de la república.*

## 4.2 Clasificación de textos: Predicción de la región de origen de un mensaje de texto

La clasificación de texto es una de las tareas emblemáticas de los modelos BERT. El problema se plantea como dado un corpus etiquetado, el modelo debe ser capaz de aprender de dicho corpus para etiquetar mensajes nunca vistos con criterios similares a los que llevaron a las etiquetas originales.

De los mensajes nunca vistos en el entrenamiento de los modelos, se creó una tarea de clasificación para medir el desempeño de dichos modelos pre-entrenados. La tarea es sobre predicción de la región de origen de mensajes confinados en el territorio nacional. Para esto se seleccionaron 313,525 mensajes geolocalizados a las 32 entidades federativas mexicanas; se cuidó que estos mensajes tuvieran información regionalizada adicional a la localidad de origen<sup>7</sup>. La colección se partió en dos subconjuntos de manera estratificada (manteniendo la proporción de la población en las clases de ambos subconjuntos) además de una proporción 70-30 para entrenamiento y prueba, respectivamente. Este procedimiento creó un conjunto de entrenamiento de 219,467 mensajes y 94,058 para prueba<sup>8</sup>. La distribución de las regiones se lista en la Tabla 3.

*Tabla 3: Distribución de los datos (tanto para entrenamiento como para prueba), ordenados de mayor a menor porcentaje (3 columnas).*

Estado	%	Estado	%	Estado	%
Ciudad de México	24.5	Yucatán	2.4	Guerrero	1.2
Nuevo León	8.2	Baja California	2.2	Tabasco	1.0
Estado de México	7.9	Zacatecas	2.0	Hidalgo	1.0
Jalisco	7.9	Sinaloa	1.9	Durango	0.8

<sup>7</sup> Mediante el clasificador MX Small sin afinar, las predicciones de los mensajes seleccionados mantenían un nivel de decisión por debajo de un umbral de entropía que permite decidir entre las diferentes regiones.

<sup>8</sup> El conjunto de datos se puede acceder desde la siguiente liga usando como clave *ingetec* <https://github.com/INGEOTEC/regional-spanish-models/raw/main/data/MX/datos-regionales-MX.zip>

Estado	%	Estado	%	Estado	%
Puebla	4.9	Morelos	1.7	Baja California Sur	0.8
Quintana Roo	3.9	Oaxaca	1.7	Aguascalientes	0.8
Veracruz	3.9	Coahuila	1.7	Campeche	0.6
Querétaro	3.5	Chiapas	1.5	Tlaxcala	0.6
Guanajuato	3.0	San Luis Potosí	1.4	Nayarit	0.5
Tamaulipas	2.7	Chihuahua	1.3	Colima	0.4
Sonora	2.7	Michoacán	1.3		

La comparación consiste en el uso de modelos pre-entrenados BETO, MX Small y MX Large, estos dos últimos corresponden a los modelos regionalizados propuestos. Adicionalmente, se añadió un modelo basado en Bolsa de Palabras (BoW) donde las palabras fueron pesadas con el modelo vectorial TFIDF<sup>9</sup> y el clasificador consiste de una máquina de soporte de vectores (SVM) con kernel lineal (usando one-vs-the-rest para manejar el problema multiclase).

La Tabla 3 muestra el desempeño en términos de la función *accuracy*<sup>10</sup> en el conjunto de prueba, con modelos afinados y sin afinar. Los modelos afinados son sometidos a un proceso de optimización adicional que permite a la red neuronal mejorar ajustando los pesos y las capas de clasificación con respecto al problema siendo resuelto. En el caso de los modelos MX Small y MX Large, no se modificó el modelo original y por lo tanto no se agregaron parámetros nuevos.

Como comparación, se usó el modelo BETO tomado de *Hugging Face*, al cual se le entrenó agregando una capa de clasificación para predecir el estado.

*Tabla 3 : Resultados en el conjunto de evaluación.*

<sup>9</sup> Frecuencia de término multiplicada por la frecuencia inversa de documento.

<sup>10</sup> Promedio de las mediciones del desempeño de las predicciones teniendo en cuenta que los aciertos de clasificación son representados como 1 y los errores como 0.

Modelo	Accuracy	
	Modelo sin afinar	Modelo afinado
BETO	-	0.6740
MX Small	0.6620	0.6970
MX Large	0.6684	<b>0.7077</b>
BoW (TFIDF+SVM)	<b>0.6747</b>	-

De los resultados podemos observar que el mejor modelo es el MX Large que obtuvo un 70.77% de aciertos a lo largo de las 32 clases. La columna *Modelo sin afinar* indica que los modelos fueron evaluados en el conjunto de prueba sin haber sido modificados de ninguna manera. El modelo BoW implica la construcción de un vocabulario (palabras), pesado de las mismas utilizando TFIDF, así como la construcción de un modelo de SVM lineal. El BoW es un modelo que puede usarse como punto de referencia ya que es ampliamente conocido, por lo mismo también da una idea de la complejidad de la tarea misma.

## 5. CONCLUSIONES

En este trabajo, presentamos un nuevo modelo de lenguaje específicamente diseñado para capturar las particularidades regionales del español mexicano. A través del entrenamiento con un corpus masivo de tuits, hemos desarrollado dos modelos con diferentes capacidades paramétricas que han demostrado ser altamente efectivos en la tarea de geolocalización de mensajes. Los resultados obtenidos en la evaluación son prometedores, ya que nuestros modelos superan al modelo de referencia BETO en la clasificación de mensajes por estado así como a modelos basados en bolsas de palabras. Esto indica que los modelos propuestos han logrado aprender las características lingüísticas distintivas de las diferentes regiones de México, lo cual representa un avance significativo en el campo del procesamiento del lenguaje natural para el español mexicano.

Al hacer públicos nuestros modelos en la plataforma Hugging Face, facilitamos su acceso a la comunidad de investigación y desarrollo, lo que permitirá impulsar futuras investigaciones en diversas áreas del Procesamiento de Lenguaje Natural, y en particular, sobre tareas y aplicaciones que deben atacarse con conocimiento regional.

## AGRADECIMIENTOS

Agradecemos el apoyo brindado a este trabajo, por el proyecto "Supercómputo como impulsor de colaboraciones academia-industria" otorgado a CIMAT por IDEA-GTO.

## 6. REFERENCIAS

- [1] T. Mikolov, K. Chen, G. Corrado, J. Dean. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781
- [2] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov (2017). Enriching word vectors with subword information. Transactions of the association for computational linguistics, 5, 135-146.
- [3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, and L. Kaiser, I. Polosukhin (2017). Attention is All you Need. Advances in Neural Information Processing Systems.
- [4] J. Devlin, M.W. Chang, K. Lee, K. Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- [5] X. Patrick, W. Shijie, V.D. Benjamin (2020). Which BERT? A Survey Organizing Contextualized Encoders. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics.
- [6] J. Cañete, G. Chaperon, R. Fuentes, J.H. Ho, H. Kang, & J. Pérez (2020). Spanish pre-trained bert model and evaluation data. PML4DC at ICLR.
- [7] E.S. Tellez, D. Moctezuma, S. Miranda, M. Graff, G. Ruiz (2023). Regionalized models for Spanish language variations based on Twitter. Language Resources and Evaluation, 57(4), 1697-1727.

- [8] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

# OPTIMIZACIÓN GLOBAL DE CLUSTERES DE BORO Y DE PLATINO DOPADOS CON METALES DE TRANSICIÓN

**Peter L. Rodríguez-Kessler**

*Centro de Investigaciones en  
Óptica A. C. (CIO)  
Loma del Bosque 115, Col. Lomas del Campestre  
C.P. 37150 León, Gto. México.  
Email: [plkessler@cio.mx](mailto:plkessler@cio.mx),  
web page: <http://www.cio.mx>*

**Alvaro Muñoz-Castro**

*Facultad de Ingeniería, Arquitectura y  
Diseño, Universidad San Sebastián,  
Bellavista 7, Santiago,  
8420524, Chile.  
Email: [alvaro.munozc@uss.cl](mailto:alvaro.munozc@uss.cl)  
web page: <http://www.amclab.cl>*

**Resumen.** La predicción de la estructura de menor energía de clústeres de boro y de metales de transición es fundamental en la ciencia de materiales. Los recientes avances en los métodos de búsqueda han reducido el tiempo de cómputo necesario para la predicción de mínimos globales. Sin embargo, el rendimiento de los resultados sigue dependiendo de la precisión de los cálculos de teoría del funcional de la densidad (DFT). En este capítulo se resumen los avances en la identificación de mínimos globales de clústeres de boro y platino dopados con metales de transición. Proporcionamos los fundamentos de los métodos de predicción de estructuras de clústeres, ejemplos de estudios recientes y perspectivas.

**Palabras clave:** Clústeres, boro, platino, DFT.

## 1 INTRODUCCIÓN

Los clústeres metálicos compuestos por un número finito de átomos han sido ampliamente estudiados, ya que se pueden considerar como un tipo de sistema intermedio entre moléculas individuales y materiales macroscópicos. En tamaños de pequeños, los clústeres y nanopartículas poseen propiedades fisicoquímicas que difieren fuertemente del bulto, y son dependientes de la forma, tamaño y composición.<sup>1-9</sup> Por lo tanto, caracterizar las estructuras favorables es una tarea importante para una comprensión más profunda y una racionalización de la química de clústeres y aleaciones. Los clústeres de metales de transición proporcionan una plataforma ideal para determinar



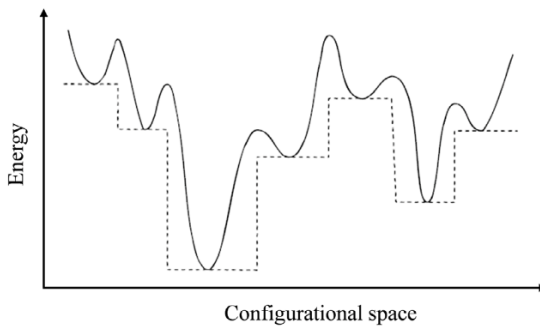
propiedades dependientes del tamaño, permitiendo establecer relaciones estructura-propiedad para una mejor orientación en la mejora de la actividad de las nanopartículas en aplicaciones deseadas.<sup>10,11</sup> En las últimas décadas, se ha llevado a cabo una extensa exploración de la superficie de energía potencial (PES) de clústeres de elementos metálicos y semiconductores, utilizando varias estrategias con el objetivo de predecir las estructuras más viables y sus propiedades de acuerdo con sus respectivos tamaños.<sup>12-18</sup> Esta tarea no es trivial, ya que el número de posibles mínimos locales en la superficie de potencial aumenta exponencialmente con el tamaño del clúster. Además de las búsquedas estructurales, los enfoques más avanzados en la actualidad revelan el gran potencial en el estudio de la dinámica de clústeres, mecanismos de dopaje, la rigidez del clúster según la temperatura, y la estabilidad termodinámica de los clústeres, entre otras características, destacando los avances recientes.<sup>19-23</sup> En este capítulo, presentamos nuestros estudios recientes y perspectivas de clústeres de boro y de platino dopados con metales de transición. Debido a la extensa literatura sobre este tema, nos centramos en ejemplos específicos. Proporcionamos una discusión basada principalmente en clústeres de  $B_7Ni_2$ ,  $B_7Cr_2$ , y  $Pt_5V$ , con el objetivo de resumir ejemplos de clústeres con posibles aplicaciones tecnológicas.

## 2 MÉTODOS DE BÚSQUEDA

La predicción de las estructuras de menor energía de los cúmulos y nanopartículas sigue siendo un problema fundamental en la ciencia de materiales. A pesar de los avances en la caracterización experimental, la forma y estructura de los clústeres no pueden determinarse directamente a partir de los experimentos. La predicción computacional de estructuras no solo sirve para contrastar y guiar los resultados experimentales, sino también para el descubrimiento de nuevos materiales nanoestructurados.<sup>24,25</sup>

La estructura de mínimo global (GM) de los clústeres metálicos generalmente se determina mediante la técnica de optimización global, que se guía mediante un gran número de optimizaciones locales.<sup>26,27</sup> Esto no es una tarea trivial, ya que el número de mínimos locales posibles en la superficie de potencial aumenta de forma exponencial con el tamaño del clúster. Por ejemplo, se han predicho al menos 1506 mínimos locales distintos para el cúmulo de 13 átomos de Lennard-Jones (LJ) y probablemente entre  $10^4$  y  $10^6$  para  $15 < n < 25$ . Para  $n = 147$ , se han sugerido alrededor de  $10^{60}$  mínimos locales.<sup>28-30</sup> Claramente, no es práctico realizar búsquedas de todos los mínimos locales de la superficie de potencial, excepto para los clústeres más pequeños.<sup>31,32</sup>

En las últimas décadas se han propuesto varios algoritmos de búsqueda, por ejemplo, el recocido simulado (SA),<sup>33</sup> el método de búsqueda aleatoria,<sup>34</sup> el basin-hopping (BH),<sup>30,35</sup> el algoritmo genético de gradiente (GEGA),<sup>36</sup> la búsqueda aleatoria de Lloyd y Johnston,<sup>37</sup> y la optimización por enjambre de partículas (PSO).<sup>38–40</sup> Más recientemente, se reportaron el método de optimización por enjambre modificado, el autómata celular y el basin-hopping Monte Carlo mejorado (BHMC), el método de crecimiento de clústeres sistemáticos (SCG),<sup>41</sup> entre otros. En su reciente trabajo, Gao et al.<sup>42</sup> proporcionaron una revisión sobre la búsqueda de estructuras inteligentes y el diseño de aplicaciones para clústeres, incluyendo los conceptos básicos y los diagramas de flujo de algunos métodos representativos como el basin-hopping,<sup>30</sup> el algoritmo evolutivo,<sup>43</sup> USPEX,<sup>44</sup> MAGUS,<sup>45</sup> SSW,<sup>46</sup> y CALYPSO.<sup>47</sup>



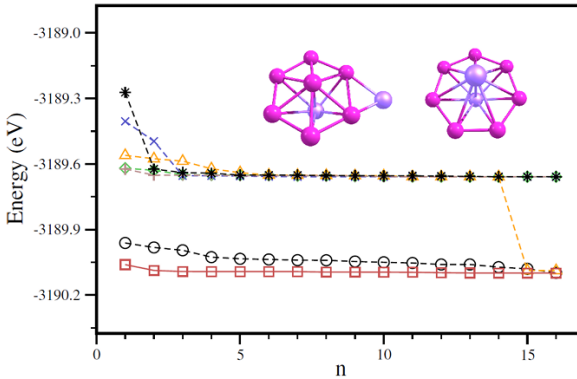
**Figura 1** Diagrama de la simulación de la superficie de potencial (PES) con el método BH. La línea continua representa la PES ideal y la línea discontinua es la PES obtenida de la simulación BH. Reproducido con permiso de la referencia [30].

Otra estrategia prometedora para resolver clústeres de mayor tamaño es combinar métodos de búsqueda de estructuras con algoritmos heurísticos como GA, BH, SA, SSW, etc., lo que ha llevado a muestras más representativas para la PES. Para cálculos precisos utilizando métodos de primeros principios como DFT, se vuelve muy difícil tratar con grandes clústeres metálicos debido al aumento exponencial del número de mínimos locales con el tamaño,<sup>35,48</sup> y la escalabilidad no lineal del costo computacional (aproximadamente  $\sim n^3$  para DFT).<sup>49</sup> Por lo tanto, se desean métodos de optimización global que requieran pocas evaluaciones de energía para su aplicación en cálculos de primeros principios. Un método prometedor para reducir el costo computacional de la evaluación de energía con DFT es utilizar

un ajuste mediante redes neuronales profundas (DNN), que pueden emplearse para evaluar la optimización local. Por lo tanto, combinar técnicas de optimización con algoritmos heurísticos y redes neuronales podría mejorar aún más el estado actual del arte en los métodos de búsqueda de estructuras. A continuación, mencionamos nuestros recientes avances en la predicción de mínimos globales de clústeres, mencionamos casos representativos incluyendo sus propiedades de estructura y estabilidad.

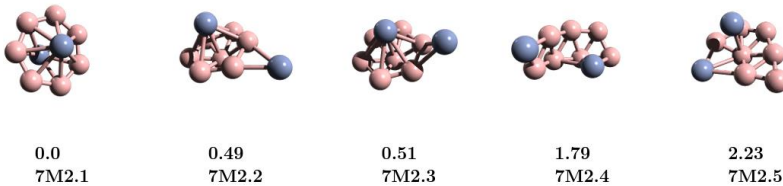
### 3 CLUSTERS DE BORO

Recientemente, se ha prestado especial atención a los clústeres de boro debido a sus aplicaciones prometedoras en biomedicina, biosensores y sistemas de liberación de fármacos.<sup>50</sup> Los clústeres de B dopados con M muestran estructuras altamente coordinadas que pueden servir como bloques de construcción para el ensamble de materiales nanoestructurados. En los últimos años se ha logrado un progreso significativo en los estudios experimentales y teóricos de clústeres de boro.<sup>51</sup> Por ejemplo, utilizando CALYPSO, se investigaron exhaustivamente las estructuras de clústeres de boro dopados con metales de transición ( $B_nM$ ).<sup>52-58</sup> Los clústeres de boro dopados se han evaluado obteniendo espectros fotoelectrónicos bien definidos, los cuales se utilizan para comprender las estructuras y los tipos de enlaces entre el boro y los metales de transición.<sup>59</sup> Al presente los estudios de clústeres de boro dopados con metales de transición  $B_nM$  ( $M=Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn$ ) son amplios en la literatura.<sup>60-63</sup> Por otro lado, la investigación sobre clústeres de boro dopados con dos átomos de metales de transición sigue siendo relativamente limitada.<sup>64</sup> Al considerar los metales de transición de la primera serie de la tabla periódica ( $TM=Fe, Co, Ni$ ), se han identificado estructuras bi-piramidales para clústeres de  $B_7M_2$ . Recientemente, hemos explorado la estructura de mínima energía del cúmulo de  $Ni_2B_7$ , encontrando patrones similares a los reportados para  $Fe_2B_7$  y  $Co_2B_7$  (ver Figura 2).<sup>65</sup> Además, estos clústeres mostraron propiedades plausibles de almacenamiento reversible de hidrógeno.



**Figura 2** Simulaciones de BH para clústeres de  $Ni_2B_7$ . Reproducido con permiso de la referencia [65].

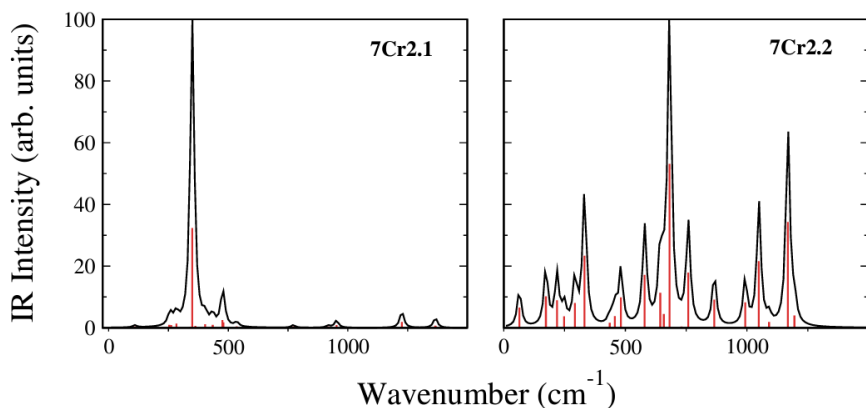
Dadas las propiedades potenciales de los clústeres de boro dopados con dos metales de transición, hemos realizado búsquedas estructurales para investigar las estructuras más estables de los clústeres de  $B_7Cr_2$ .<sup>64</sup> Las estructuras fueron obtenidas utilizando un método de basin-hopping modificado a lo largo de siete generaciones, como se discute en trabajos previos.<sup>41</sup> La Figura 3 muestra la estructura más estable y los isómeros representativos, en orden creciente de energía. Los resultados mostraron que la estructura más estable tiene una configuración bi-piramidal, o estructura de sándwich inverso. Los isómeros de mayor energía, adoptan estructuras de menor simetría, con segregación de los metales de transición hacia los sitios periféricos.



**Figura 3** Estructuras de menor energía para los clústeres de  $B_7Cr_2$ . Para cada estructura, se indican la energía relativa (en eV) y la etiqueta del isómero. Reproducido con permiso de la referencia [64].

Al evaluar las propiedades de los clústeres, se encontraron diferencias notables entre el mínimo global y los isómeros de mayor energía. Por ejemplo, se calcularon los espectros IR de los dos clústeres más estables, los cuales pueden servir como guía para futuros experimentos. Se encontraron picos característicos para los clústeres **7M2.1** y **7M2.2** en  $349.35$  y  $681.88$   $cm^{-1}$ ,

respectivamente, sugiriendo diferencias notables en sus estructuras. Los espectros IR de los clústeres muestran una dispersión significativa para **7M2.2**, mientras son más localizados para el mínimo global **7M2.1** (Figura 3).



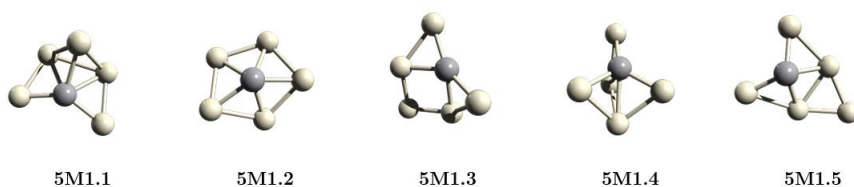
*Figura 3* Espectros IR para los clústeres **7M2.1** y **7M2.2** obtenidos a nivel PBE0/Def2TZVP. Reproducido con permiso de la referencia<sup>[64]</sup>.

Las frecuencias vibracionales más bajas se encuentran en 109.18 y 64.20  $\text{cm}^{-1}$ , mientras que las frecuencias más altas están en 1366.26 y 1195.07  $\text{cm}^{-1}$ , respectivamente. Estos datos sirven como base para futuras investigaciones en aplicaciones específicas de estos clústeres. Además, los espectros IR pueden ser contrastados con futuros experimentos.

## 4 CLUSTERES DE PLATINO

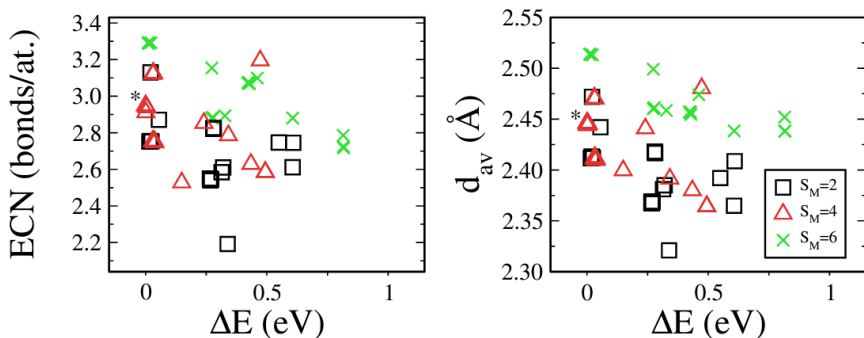
Los nanoclústeres basados en platino han sido ampliamente investigados debido a sus propiedades potenciales en catálisis heterogénea. La investigación en materiales de platino ha llevado a la evaluación de sitios activos catalíticos, permitiendo estrategias adicionales hacia materiales catalíticos más eficientes. En particular, las aleaciones de platino con metales no-preciosos (M) ofrecen reducciones en el costo, así como un aumento en las actividades catalíticas.<sup>66,67</sup> Para los clústeres de platino puro, las estructuras para tamaños pequeños se encuentran bien identificadas, mientras que para una atomicidad de 16 en adelante, las estructuras del mínimo global aún se encuentran en debate.<sup>68,69</sup> Por otro lado, los estudios sobre clústeres de Pt dopados con los primeros metales de transición son limitados.<sup>70</sup> Esto no es una tarea trivial, ya que los clústeres dopados requieren una exploración extensa de sus estructuras, orden químico y espín.<sup>71</sup> En la literatura, Jennings

y Johnston revelaron las estructuras más estables de los cúmulos de  $Pt_{x-y}M_y$  (con  $x=2-6$ ,  $y=1,2$  y  $M=Ti, V$ ) y encontraron que los estados de espín tienen un efecto despreciable en las estructuras de los cúmulos.<sup>72</sup> Con el fin de confirmar la estructura de menor energía de estos cúmulos, en nuestro reciente trabajo, probamos el caso de los cúmulos de  $Pt_5V$  utilizando varias funcionales.<sup>73</sup>



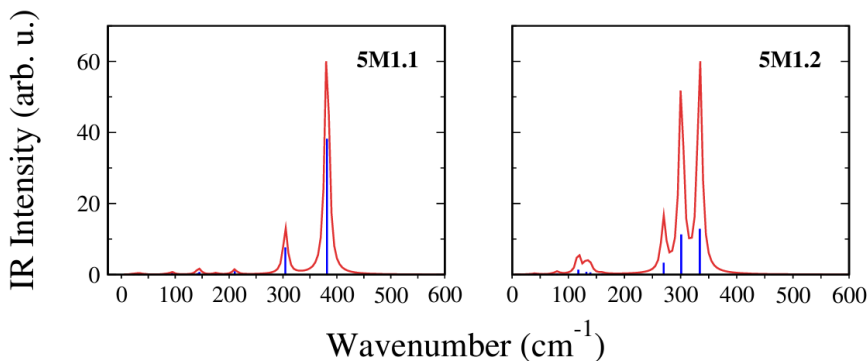
**Figura 4** Las estructuras de menor energía de los cúmulos de  $Pt_5V$ . Las estructuras están etiquetadas con la notación  $nMm.y$ , donde  $n$  representa el número de átomos de platino,  $M$  denota el sistema dopado, y  $y$  se refiere al número de isómero en orden creciente de energía. Reproducido con permiso de la referencia [73].

Los resultados muestran que la estructura de mínimo global exhibe un comportamiento dependiente de la funcional. Considerando las funcionales híbridas como PBE0 y B3PW91, se favorece la estabilidad del isómero **5M1.1**, mientras que considerando las funcionales de tipo meta-GGA como TPSS y M06-L se favorece la estabilidad de la estructura **5M1.2** que coincide con el mínimo global reportado por Jennings y Johnson. Estos resultados sugieren que posiblemente las propiedades de estos sistemas no dependan solo de una estructura sino de los isómeros de más baja energía. Después de analizar los resultados en varias funcionales, discutimos además las propiedades de las estructuras más viables a nivel B3PW91/Def2TZVP, siendo un enfoque razonable para describir las longitudes de enlace y energías de ionización de los cúmulos de platino.<sup>74</sup> Las propiedades estructurales se evalúan además utilizando los parámetros del número de coordinación efectivo (ECN) y la distancia de enlace promedio ( $d_{av}$ ).<sup>75</sup> Los resultados muestran que el ECN aumenta con la multiplicidad de espín ( $S_M$ ), lo que indica una dependencia de la estructura y el carácter magnético de los clusters (Figura 5).



**Figura 5** Distribución del número de coordinación efectivo y la distancia de enlace promedio en función de la energía relativa ( $\Delta E$ ) de los cúmulos de  $Pt_5V$  evaluados a nivel  $B3PW91/Def2TZVP$ . Los valores de las multiplicidades de espín ( $S_M$ ) se indican con símbolos distintos. La configuración más estable se especifica con un (\*). Reproducido con permiso de la referencia [73].

Con el objetivo de discutir las propiedades vibracionales de los cúmulos de  $Pt_5V$  representativos, hemos calculado además los espectros infrarrojos (IR), que servirán para contrastar futuros experimentos cuando estén disponibles. Los picos característicos para **5M1.1** se encuentran en  $304.36$  y  $381.58$   $cm^{-1}$ , mientras que para **5M1.2** están en  $301.23$  y  $335.16$   $cm^{-1}$ , respectivamente (Figura 6).



**Figura 6** Espectros IR para los cúmulos **5M1.1** y **5M1.2** obtenidos a nivel  $B3PW91/Def2TZVP$ . Reproducido con permiso de la referencia [73].

Las frecuencias vibracionales más bajas y más altas para **5M1.1** se encuentran en  $22.80$ - $381.58$   $cm^{-1}$ , mientras que para **5M1.2** están en  $29.83$ - $334.16$   $cm^{-1}$ , respectivamente, lo que indica un pequeño rango de los espectros. Los parámetros de energía de ionización (IP) y afinidad electrónica (EA) son

cantidades físicas importantes que reflejan la estabilidad electrónica de los clústeres. Los resultados mostraron que los isómeros de  $Pt_5V$  exhiben solo pequeñas variaciones en las propiedades electrónicas. Para los isómeros  $5M1.y$  ( $y=1-5$ ), los parámetros de IP y EA varían entre 7.05-7.58 y 2.16-2.37 eV, respectivamente. Los parámetros de dureza química ( $\eta$ ) y potencial químico ( $\mu$ ) también mostraron variaciones mínimas entre estos isómeros representativos. En este contexto, las diferencias entre los cúmulos **5M1.1** y **5M1.2** fueron despreciables.

## 5 CONCLUSIONES Y PERSPECTIVAS

En las últimas décadas, se ha llevado a cabo una extensa exploración de la superficie de energía potencial (PES) de clústeres de semiconductores y metales de transición, utilizando varias estrategias para predecir las estructuras de mínima energía. En este capítulo, se discutieron los fundamentos, el progreso y los avances recientes en la identificación de la estructura de energía mínima de clústeres dopados con metales de transición, con especial atención a los clústeres de boro y platino. Se presenta una visión general de los avances en métodos de búsqueda de mínimos globales de clústeres y nanoestructuras. Los métodos mejorados, junto con las respectivas comparaciones con los resultados experimentales, son beneficiosos para construir una base sólida y confiable para una comprensión racional de las propiedades de los clústeres basadas en sus características dependientes del tamaño. En este capítulo, hemos utilizado la teoría del funcional de la densidad (DFT) para explorar la estructura de los cúmulos de  $B_7Ni_2$ ,  $B_7Cr_2$ , y  $Pt_5V$ , respectivamente. Los resultados de los clusters de boro mostraron características representativas de las estructuras de mínima energía con respecto a los isómeros de mayor energía. Mientras que las estructuras de platino mostraron una distribución de los isómeros en función del espín y de la coordinación de los cúmulos. Estos isómeros mostraron propiedades electrónicas menos sensibles a la estructura que los isómeros de boro, lo que se podría explicar por los distintos entornos de los dopantes, que se presentan en los isómeros de boro y platino. Estos resultados sirven como base para futuras investigaciones de las estructuras de los clústeres de boro y platino en función del tamaño.



## 6 AGRADECIMIENTOS

P. L. R.-K. agradece los recursos de cómputo proporcionados por los Laboratorios de Supercómputo del CIMAT Campus Guanajuato y Puerto Interior, los cuales han sido fundamentales para la realización de la presente investigación. A. M.-C. quisiera agradecer el apoyo del FONDECYT ANID Regular 1221676.

## 7 REFERENCIAS

1. Doud EA, Voevodin A, Hochuli TJ, Champsaur AM, Nuckolls C, Roy X. Superatoms in materials science. *Nat Rev Mater.* 2020;5(5):371-387. doi:10.1038/s41578-019-0175-3
2. Jena P, Sun Q. Super Atomic Clusters: Design Rules and Potential for Building Blocks of Materials. *Chem Rev.* 2018;118(11):5755-5870. doi:10.1021/acs.chemrev.7b00524
3. Chakraborty I, Pradeep T. Atomically Precise Clusters of Noble Metals: Emerging Link between Atoms and Nanoparticles. *Chem Rev.* 2017;117(12):8208-8271. doi:10.1021/acs.chemrev.6b00769
4. Heider Y, Scheschkewitz D. Molecular Silicon Clusters. *Chem Rev.* 2021;121(15):9674-9718. doi:10.1021/acs.chemrev.1c00052
5. Zhao J, Du Q, Zhou S, Kumar V. Endohedrally Doped Cage Clusters. *Chem Rev.* 2020;120(17):9021-9163. doi:10.1021/acs.chemrev.9b00651
6. Luo Z, Castleman AW, Khanna SN. Reactivity of Metal Clusters. *Chem Rev.* 2016;116(23):14456-14492. doi:10.1021/acs.chemrev.6b00230
7. Dash BP, Satapathy R, Maguire JA, Hosmane NS. Polyhedral boron clusters in materials science. *New J Chem.* 2011;35(10):1955. doi:10.1039/c1nj20228f
8. Pan F, Peerless B, Dehnen S. Bismuth-Based Metal Clusters—From Molecular Aesthetics to Contemporary Materials Science. *Acc Chem Res.* 2023;56(9):1018-1030. doi:10.1021/acs.accounts.3c00020
9. Pan F, Guggolz L, Dehnen S. Cluster Chemistry with ( Pseudo - )Tetrahedra Involving Group 13–15 (Semi-)Metal Atoms. *CCS Chem.* 2022;4(3):809-824. doi:10.31635/ccschem.021.202101340
10. Ferrando R, Jellinek J, Johnston RL. Nanoalloys: From Theory to Applications of Alloy Clusters and Nanoparticles. *Chem Rev.* 2008;108(3):845-910. doi:10.1021/cr040090g
11. Daniel MC, Astruc D. Gold nanoparticles: assembly, supramolecular chemistry, quantum-size-related properties, and applications

toward biology, catalysis, and nanotechnology. *Chem Rev.* 2004;104(1):293-346. doi:10.1021/cr030698+

12. Sun G, Sautet P. Metastable Structures in Cluster Catalysis from First-Principles: Structural Ensemble in Reaction Conditions and Metastability Triggered Reactivity. *J Am Chem Soc.* 2018;140(8):2812-2820. doi:10.1021/jacs.7b11239

13. Ballone P. Modeling Potential Energy Surfaces: From First-Principle Approaches to Empirical Force Fields. *Entropy.* 2013;16(1):322-349. doi:10.3390/e16010322

14. Jana G, Pal R. Generation of global minimum energy structures of small molecular clusters using machine learning technique. In: *Atomic Clusters with Unusual Structure, Bonding and Reactivity.* Elsevier; 2023:185-212. doi:10.1016/B978-0-12-822943-9.00001-2

15. Marques JMC, Pereira FB, Llanio-Trujillo JL, et al. A global optimization perspective on molecular clusters. *Philos Trans R Soc A Math Phys Eng Sci.* 2017;375(2092):20160198. doi:10.1098/rsta.2016.0198

16. Roth M, Toker Y, Major DT. Monte Carlo-Simulated Annealing and Machine Learning-Based Funneled Approach for Finding the Global Minimum Structure of Molecular Clusters. *ACS Omega.* 2024;9(1):1298-1309. doi:10.1021/acsomega.3c07600

17. Zhang J, Dolg M. Global optimization of clusters of rigid molecules using the artificial bee colony algorithm. *Phys Chem Chem Phys.* 2016;18(4):3003-3010. doi:10.1039/C5CP06313B

18. Yañez O, Báez-Grez R, Inostroza D, et al. AUTOMATON: A Program That Combines a Probabilistic Cellular Automata and a Genetic Algorithm for Global Minimum Search of Clusters and Molecules. *J Chem Theory Comput.* 2019;15(2):1463-1475. doi:10.1021/acs.jctc.8b00772

19. Sabooni Asre Hazer M, Malola S, Häkkinen H. Isomer dynamics of the  $[\text{Au}_6(\text{NHC-S})_4]^{2+}$  nanocluster. *Chem Commun.* 2022;58(19):3218-3221. doi:10.1039/D2CC00676F

20. Malola S, Häkkinen H. Prospects and challenges for computer simulations of monolayer-protected metal clusters. *Nat Commun.* 2021;12(1):2197. doi:10.1038/s41467-021-22545-x

21. Malola S, Matus MF, Häkkinen H. Theoretical Analysis of the Electronic Structure and Optical Properties of DNA-Stabilized Silver Cluster  $\text{Ag}_{16}\text{Cl}_2$  in Aqueous Solvent. *J Phys Chem C.* 2023;127(33):16553-16559. doi:10.1021/acs.jpcc.3c04103

22. Juarez-Mosqueda R, Malola S, Häkkinen H. Ab initio molecular dynamics studies of  $\text{Au}_{38}(\text{SR})_{24}$  isomers under heating. *Eur Phys J D.*

2019;73(3):62. doi:10.1140/epjd/e2019-90441-5

23. Kacprzak KA, Akola J, Häkkinen H. First-principles simulations of hydrogen peroxide formation catalyzed by small neutral gold clusters. *Phys Chem Chem Phys*. 2009;11(30):6359. doi:10.1039/b822765a

24. Chaves AS, Piotrowski MJ, Da Silva JLF. Evolution of the structural, energetic, and electronic properties of the 3d, 4d, and 5d transition-metal clusters (30 TM n systems for n = 2–15): a density functional theory investigation. *Phys Chem Chem Phys*. 2017;19(23):15484-15502. doi:10.1039/C7CP02240A

25. Bulusu S, Yoo S, Zeng XC. Search for global minimum geometries for medium sized germanium clusters: Ge12–Ge20. *J Chem Phys*. 2005;122(16). doi:10.1063/1.1883647

26. Zhao Y, Chen X, Li J. TGMIn: A global-minimum structure search program based on a constrained basin-hopping algorithm. *Nano Res*. 2017;10(10):3407-3420. doi:10.1007/s12274-017-1553-z

27. Hartke B. Global optimization. *WIREs Comput Mol Sci*. 2011;1(6):879-887. doi:10.1002/wcms.70

28. Tsai CJ, Jordan KD. Use of an eigenmode method to locate the stationary points on the potential energy surfaces of selected argon and water clusters. *J Phys Chem*. 1993;97(43):11227-11237. doi:10.1021/j100145a019

29. Cheng J, Fournier R. Structural optimization of atomic clusters by tabu search in descriptor space. *Theor Chem Accounts Theory, Comput Model (Theoretica Chim Acta)*. 2004;112(1):7-15. doi:10.1007/s00214-003-0552-1

30. Wales DJ, Doye JPK. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J Phys Chem A*. 1997;101(28):5111-5116. doi:10.1021/jp970984n

31. Northby JA. Structure and binding of Lennard-Jones clusters:  $13 \leq N \leq 147$ . *J Chem Phys*. 1987;87(10):6166-6177. doi:10.1063/1.453492

32. Xue G. Improvement on the northby algorithm for molecular conformation: Better solutions. *J Glob Optim*. 1994;4(4):425-440. doi:10.1007/BF01099267

33. No Title. doi:10.1126/science.220.4598.671

34. Saunders M. Stochastic search for isomers on a quantum mechanical surface. *J Comput Chem*. 2004;25(5):621-626. doi:10.1002/jcc.10407

35. Heiles S, Johnston RL. Global optimization of clusters using electronic structure methods. *Int J Quantum Chem*. 2013;113(18):2091-2109. doi:10.1002/qua.24462

36. Alexandrova AN. H·(H<sub>2</sub>O)<sub>n</sub> Clusters: Microsolvation of the Hydrogen Atom via Molecular ab Initio Gradient Embedded Genetic

- Algorithm (GEGA). *J Phys Chem A*. 2010;114(48):12591-12599.  
doi:10.1021/jp1092543
37. Lloyd LD, Johnston RL. Modelling aluminium clusters with an empirical many-body potential. *Chem Phys*. 1998;236(1-3):107-121.  
doi:10.1016/S0301-0104(98)00180-3
38. Jana G, Mitra A, Pan S, Sural S, Chattaraj PK. Modified Particle Swarm Optimization Algorithms for the Generation of Stable Structures of Carbon Clusters,  $C_n$  ( $n = 3-6, 10$ ). *Front Chem*. 2019;7.  
doi:10.3389/fchem.2019.00485
39. Call ST, Zubarev DY, Boldyrev AI. Global minimum structure searches via particle swarm optimization. *J Comput Chem*. 2007;28(7):1177-1186. doi:10.1002/jcc.20621
40. Tong Q, Lv J, Gao P, Wang Y. The CALYPSO methodology for structure prediction\*. *Chinese Phys B*. 2019;28(10):106105. doi:10.1088/1674-1056/ab4174
41. Rodríguez-Kessler PL, Rodríguez-Domínguez AR, Muñoz-Castro A. Systematic cluster growth: a structure search method for transition metal clusters. *Phys Chem Chem Phys*. 2021;23(8):4935-4943.  
doi:10.1039/D0CP06179D
42. Gao J, Zhao L, Chang Y, et al. Intelligent Structure Searching and Designs for Nanoclusters: Effective Units in Atomic Manufacturing. *Adv Intell Syst*. Published online April 26, 2024. doi:10.1002/aisy.202300716
43. Katoch S, Chauhan SS, Kumar V. A review on genetic algorithm: past, present, and future. *Multimed Tools Appl*. 2021;80(5):8091-8126.  
doi:10.1007/s11042-020-10139-6
44. Oganov AR, Glass CW. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *J Chem Phys*. 2006;124(24). doi:10.1063/1.2210932
45. Wang J, Gao H, Han Y, et al. MAGUS: machine learning and graph theory assisted universal structure searcher. *Natl Sci Rev*. 2023;10(7).  
doi:10.1093/nsr/nwad128
46. Shang C, Liu ZP. Stochastic Surface Walking Method for Structure Prediction and Pathway Searching. *J Chem Theory Comput*. 2013;9(3):1838-1845. doi:10.1021/ct301010b
47. Wang H, Wang Y, Lv J, Li Q, Zhang L, Ma Y. CALYPSO structure prediction method and its wide application. *Comput Mater Sci*. 2016;112:406-415. doi:10.1016/j.commatsci.2015.09.037
48. Pickard CJ, Needs RJ. Ab initio random structure searching. *J Phys Condens Matter*. 2011;23(5):053201. doi:10.1088/0953-8984/23/5/053201

49. Ouyang R, Xie Y, Jiang D en. Global minimization of gold clusters by combining neural network potentials and the basin-hopping method. *Nanoscale*. 2015;7(36):14817-14821. doi:10.1039/C5NR03903G
50. Shakerzadeh E, Tahmasebi E, Van Duong L, Tho Nguyen M. Boron Clusters in Biomedical Applications: A Theoretical Viewpoint. In: *Characteristics and Applications of Boron*. IntechOpen; 2022. doi:10.5772/intechopen.106215
51. Kiran B, Bulusu S, Zhai HJ, Yoo S, Zeng XC, Wang LS. Planar-to-tubular structural transition in boron clusters: B<sub>20</sub> as the embryo of single-walled boron nanotubes. *Proc Natl Acad Sci*. 2005;102(4):961-964. doi:10.1073/pnas.0408132102
52. Lv J, Wang Y, Zhang L, Lin H, Zhao J, Ma Y. Stabilization of fullerene-like boron cages by transition metal encapsulation. *Nanoscale*. 2015;7(23):10482-10489. doi:10.1039/C5NR01659B
53. Chen B Le, Sun WG, Kuang XY, et al. Structural Stability and Evolution of Medium-Sized Tantalum-Doped Boron Clusters: A Half-Sandwich-Structured TaB<sub>12</sub> – Cluster. *Inorg Chem*. 2018;57(1):343-350. doi:10.1021/acs.inorgchem.7b02585
54. Chen B, Sun W, Kuang X, et al. Insights into the effects produced by doping of medium-sized boron clusters with ruthenium. *Phys Chem Chem Phys*. 2018;20(48):30376-30383. doi:10.1039/C8CP05725G
55. Li P, Du X, Wang JJ, Lu C, Chen H. Probing the Structural Evolution and Stabilities of Medium-Sized MoB<sub>n</sub><sup>0/-</sup> Clusters. *J Phys Chem C*. 2018;122(34):20000-20005. doi:10.1021/acs.jpcc.8b05759
56. Sun W, Xia X, Lu C, Kuang X, Hermann A. Probing the structural and electronic properties of zirconium doped boron clusters: Zr distorted B<sub>12</sub> ligand framework. *Phys Chem Chem Phys*. 2018;20(36):23740-23746. doi:10.1039/C8CP03384F
57. Jin S, Chen B, Kuang X, Lu C, Gutsev GL. Structural evolution and electronic properties of medium-sized boron clusters doped with scandium. *J Phys Condens Matter*. 2019;31(48):485302. doi:10.1088/1361-648X/ab3b2b
58. Shakerzadeh E, Van Duong L, Tahmasebi E, Nguyen MT. The scandium doped boron cluster B<sub>27</sub> Sc<sub>2</sub><sup>+</sup>: a fruit can-like structure. *Phys Chem Chem Phys*. 2019;21(17):8933-8939. doi:10.1039/C9CP00892F
59. Chen WJ, Kulichenko M, Choi HW, et al. Photoelectron Spectroscopy of Size-Selected Bismuth–Boron Clusters: BiB<sub>n</sub> – (n = 6–8). *J Phys Chem A*. 2021;125(31):6751-6760. doi:10.1021/acs.jpca.1c05846
60. Li SX, Yang YJ, Chen DL, Long ZW. Structures, and electronic and spectral properties of single-atom transition metal-doped boron clusters MB<sub>24</sub> – (M = Sc, Ti, V, Cr, Mn, Fe, Co, and Ni). *RSC Adv*. 2022;12(26):16706-16716. doi:10.1039/D2RA02500K

61. Ren RQ, Long ZW, Li SX, Gao CG. Size effects and electronic properties of zinc-doped boron clusters  $Zn_n$  ( $n = 1-15$ ). *J Mol Model*. 2024;30(5):123. doi:10.1007/s00894-024-05906-3
62. Rodríguez-Kessler PL, Vásquez-Espinal A, Muñoz-Castro A. Structure and stability of Cu-doped B ( $n = 1-12$ ) clusters: DFT calculations. *Polyhedron*. 2023;243:116538. doi:10.1016/j.poly.2023.116538
63. Rodríguez-Kessler PL, Vásquez-Espinal A, Rodríguez-Domínguez AR, Cabellos-Quiroz JL, Muñoz-Castro A. Structures of Ni-doped B ( $n = 1-13$ ) clusters: A computational study. *Inorganica Chim Acta*. 2024;568:122062. doi:10.1016/j.ica.2024.122062
64. Rodríguez-Kessler PL. On the structures and stabilities of  $B_7Cr_2$  clusters: A DFT study. Published online July 29, 2024. <http://arxiv.org/abs/2407.20498>
65. Olalde-López D, Rodríguez-Kessler PL, Rodríguez-Carrera S, Muñoz-Castro A. Hydrogen storage properties for bimetallic doped boron clusters  $M_2B$  ( $M = Fe, Co, Ni, Cu$ ). *Int J Hydrogen Energy*. Published online June 2024. doi:10.1016/j.ijhydene.2024.05.429
66. Stamenkovic VR, Mun BS, Arenz M, et al. Trends in electrocatalysis on extended and nanoscale Pt-bimetallic alloy surfaces. *Nat Mater*. 2007;6(3):241-247. doi:10.1038/nmat1840
67. Xiong L, Kannan A, Manthiram A. Pt-M ( $M = Fe, Co, Ni$  and Cu) electrocatalysts synthesized by an aqueous route for proton exchange membrane fuel cells. *Electrochem Commun*. 2002;4(11):898-903. doi:10.1016/S1388-2481(02)00485-X
68. Guevara-Vela JM, Rocha-Rinza T, Rodríguez-Kessler PL, Muñoz-Castro A. On the structure and electronic properties of Pt  $n$  clusters: new most stable structures for  $n = 16-17$ . *Phys Chem Chem Phys*. 2023;25(42):28835-28840. doi:10.1039/D3CP04455F
69. Guevara-Vela JM, Gallegos M, Rocha-Rinza T, Muñoz-Castro A, Kessler PLR, Martín Pendás Á. New global minimum conformers for the  $Pt_{19}$  and  $Pt_{20}$  clusters: low symmetric species featuring different active sites. *J Mol Model*. 2024;30(9):310. doi:10.1007/s00894-024-06099-5
70. Rodríguez-Kessler PL, Rodríguez-Domínguez AR, Alonso-Dávila P, Navarro-Santos P, Muñoz-Castro A. Structural and electronic properties for Be-doped Pt  $n$  ( $n = 1-12$ ) clusters obtained by DFT calculations. *Phys*

*Chem Chem Phys.* 2022;24(13):7856-7861. doi:10.1039/D1CP05410D

71. Chittari BL, Kumar V. Atomic structure, alloying behavior, and magnetism in small Fe-Pt clusters. *Phys Rev B.* 2015;92(12):125442.

doi:10.1103/PhysRevB.92.125442

72. No Title. doi:10.1016/j.comptc.2013.06.033

73. Rodríguez-Kessler PL. Revisiting the Global Minimum Structure of the Pt<sub>5</sub>V Cluster. Published online August 7, 2024.

<http://arxiv.org/abs/2408.03537>

74. Guevara-Vela JM, Rocha-Rinza T, Rodríguez-Kessler PL, Muñoz-Castro A. On the structure and electronic properties of Pt *n* clusters: new most stable structures for *n* = 16–17. *Phys Chem Chem Phys.*

2023;25(42):28835-28840. doi:10.1039/D3CP04455F

75. Rodríguez-Kessler PL, Rodríguez-Domínguez AR. Stability of Ni Clusters and the Adsorption of CH<sub>4</sub>: First-Principles Calculations. *J Phys Chem C.* 2015;119(22):12378-12384. doi:10.1021/acs.jpcc.5b01738

# CLASIFICACIÓN DE ACTIVIDADES HUMANAS ANÓMALAS EN UN ENTORNO DE VIDEO VIGILANCIA.

**Jaqueline Reynosa-Guerrero, Alberto Vázquez-Daniela Moctezuma**

*Centro de Investigación en Ciencias de Información Geoespacial. (CentroGEO)  
Circuito Tecnopolo Norte, 107, Col. Tecnopolo Pocitos II, Aguascalientes, Ags. CP 20313.  
Email: [reynosa.jaqueline@gmail.com](mailto:reynosa.jaqueline@gmail.com),  
[dmoctezuma@centrogeo.edu.mx](mailto:dmoctezuma@centrogeo.edu.mx)  
web page: <https://www.centrogeo.org.mx>*

**Cervantes.**

*Centro de Ingeniería y Desarrollo Industrial (CIDESI)  
Av pie de la Cuesta 702, Desarrollo San Pablo, Oro, Oro.  
Email: [albertovcervantes@gmail.com](mailto:albertovcervantes@gmail.com),  
web page: <https://cidesi.com>*

**Resumen.** Este trabajo surge a partir del proyecto titulado “Generación de sistemas inteligentes aplicado a temas de seguridad ciudadana” de la convocatoria de estancias posdoctorales del Conahcyt. El objetivo del proyecto es la clasificación de actividades humanas, en este caso de actividades atípicas o anormales en una secuencia de videovigilancia con propósitos de asistir en temas de seguridad ciudadana. Para tener datos con los cuales probar los modelos, se realizó una búsqueda de una base de datos que contuviera videos de actividades normales, así como con actividades anómalas, se escogió la base *UCF crime* con 13 clases de acciones, las cuales son: Abuso, Arresto, Incendio premeditado, Agresión, Robo con allanamiento, Explosión, Pelea, Accidentes de tráfico, Robo, Tiroteo, Robo en tiendas, Robo (sin allanamiento), y Vandalismo. La gran mayoría de los trabajos reportados en el estado del arte para la clasificación de estas acciones<sup>1,2,3</sup> se basan en modelos de redes neuronales profundas con arquitecturas como CNNs (Convolutional Neural Networks) y RNNs (Recurrent Neural Networks). En este trabajo se evalúa la detección de actividades anómalas de acuerdo con el trabajo realizado en “Learning Prompt-Enhanced Context Features for Weakly-Supervised Video Anomaly Detection”<sup>3</sup>, así como posibles mejoras a este.



**Palabras clave:** Videovigilancia, clasificación actividades humanas, detección de acciones anómalas, deep learning.

## INTRODUCCIÓN

En la actualidad, la seguridad ciudadana es una preocupación primordial en sociedades modernas, impulsando la necesidad de desarrollar sistemas inteligentes capaces de detectar y prevenir actividades anómalas que puedan representar un riesgo para la población. Esto lleva a la búsqueda de trabajos de inteligencia artificial relacionados con código abierto de sistemas avanzados de detección de anomalías en videos de vigilancia, para lo cual podemos utilizar la base de datos UCF crime la cual es un recurso valioso con una gran colección de videos que capturan actividades normales y actividades anómalas como asaltos, vandalismo y otros actos.

Existen numerosos estudios que abordan diferentes enfoques para identificar actividades normales y anormales en videovigilancia. Por ejemplo, un enfoque basado en redes neuronales convolucionales<sup>7</sup> profundiza en el uso de flujos ópticos y mapas de saliencia para la detección efectiva de acciones anómalas en secuencias de video. Otro enfoque utiliza técnicas de aprendizaje débilmente supervisado<sup>8</sup>, donde se combinan métodos de normalización por lotes para mejorar la precisión y eficiencia en la identificación de anomalías, lo que demuestra la viabilidad de modelos menos dependientes de la anotación manual extensa. Además, estudios recientes han propuesto el uso de modelos basados en memoria, que retienen patrones aprendidos<sup>9</sup> de actividades normales para detectar desviaciones que podrían indicar la presciencia de una anomalía. Por lo cual para realizar este reporte se tomaron los trabajos relacionados a la base de datos UCF-crime y se evaluaron algunos códigos relacionados utilizando la infraestructura de los servidores de CIMAT, El Insurgente y Puerto Interior, situados en el estado de Guanajuato.

## TRABAJOS RELACIONADOS

Uno de los trabajos relacionados en la detección de anomalías es el titulado “Localizing Anomalies from Weakly-Labeled Videos”<sup>1</sup>, donde proponen el método WSAL (*Weakly Supervised Anomaly Localization*) que se basa en la evaluación de la evolución de segmentos temporales adyacentes en videos anómalos, aprovechando las diferencias de apariencia para detectar cambios significativos que indiquen anomalías. Una de sus contribuciones es reforzar la capacidad del modelo para detectar y localizar anomalías con la creación de anomalías artificiales. Además, los autores proponen un conjunto de datos propio de condiciones anómalas de tráfico. La detección de anomalías es

igual a los valores atípicos de una escena considerada normal, sin embargo, esta técnica normalmente falla cuando se enfrenta a escenas más complicadas o desconocidas. Los autores de “Localizing Anomalies from Weakly-Labeled Videos”<sup>1</sup> comparan la precisión de localización de anomalías en términos de AUC (área bajo la curva), donde alcanzan en “Real-world anomaly detection in surveillance videos”<sup>2</sup>, alcanzan un AUC de 54.25% y en “Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection”<sup>3</sup> alcanza un 59.02%, mientras el propuesto en “Localizing Anomalies from Weakly-Labeled Videos”<sup>1</sup> es de 67.38%. Este método “Localizing Anomalies from Weakly-Labeled Videos”<sup>1</sup> primero aplica una capa convolucional para la detección de características de cada cuadro. Para predecir si un video es normal o anómalo, basado en la localización temporal de las anomalías, el método propuesto divide el video en segmentos y utiliza una función de alto orden para codificar tanto el segmento anclado como sus segmentos adyacentes en el contexto temporal.

En el escenario de supervisión débil, se da una secuencia de video  $X$  y su correspondiente anotación a nivel de video “ $y \in \{0,1\}$ ”, donde el caso ‘ $y=1$ ’ significa que existe una anomalía en esta secuencia, y ‘ $y=0$ ’ indica que no hay anomalía en  $X$ .

$S(X)$  es la puntuación de un video que calcula la distancia relativa máxima entre pares de posiciones, se espera que las puntuaciones de los videos normales sean menores que las de los videos anómalos. Por lo tanto, la estrategia de máxima distancia impone que los videos normales sean más uniformes que los videos con anomalías, lo que ajusta a la suposición convencional.

$$S(X) = \max_{i,j=1,\dots,m} f(\varphi(x_{i-k},\dots,x_{i+k}), \varphi(x_{j-k},\dots,x_{j+k})) \quad (1)$$

Donde  $f$  es una medida de distancia que mide el margen de las puntuaciones de anomalías entre las posiciones de los segmentos  $i$  y  $j$ . Cuanto más cercanas sean las puntuaciones de anomalía predichas, menor será la distancia. Las puntuaciones de anomalía están todas cerca de cero en un video normal, mientras en un video anómalo, las anomalías que se encuentran en un fondo normal se llevarán una puntuación grande.

En este mismo trabajo de “Localizing Anomalies from Weakly-Labeled Videos”<sup>1</sup> aplicando Codificación de contexto de Alto orden (High-order Context Encoding, HCE) se puede determinar las variaciones dinámicas en las

series temporales que indican la existencia de anomalías, además se introduce una anomalía generada manualmente en la secuencia para mejorar la detección de anomalías. Finalmente se utiliza ROC como métrica estándar ajustando los hiperparámetros del modelo y evaluar su capacidad de localizar anomalías.

Otro trabajo es (MGFN: Magnitude-Contrastive Glance-and-Focus Network for Weakly-Supervised Video Anomaly Detection)<sup>4</sup> que se enfoca en amplificar las características para determinar las anomalías, donde el término anomalía está definido en contraposición de lo que es normalidad, para la clasificación se presentan dos enfoques, una arquitectura espacio-temporal, que se basa en construir modelos que integran la información espacial y temporal, pero enfrentan dificultades en videos largos donde las anomalías ocupan solamente una pequeña parte del video, el otro enfoque propone una función de pérdida de Magnitud de Características Temporales (RTFM) para que las características normales sean más grandes, como se observa en la figura 1.

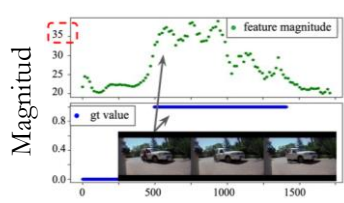


Figura 1 –Magnitud de las características con anomalías.

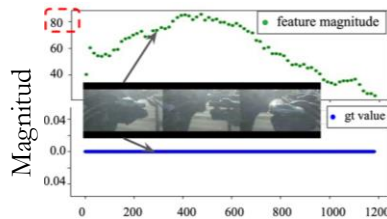


Figura 2 – Se observa una seccion de video normal aonae la magnitud de las características puede ser mayor que la magnitud de las características con anomalías, como en la figura 1.

Así que en (MGFN: Magnitude-Contrastive Glance-and-Focus Network for Weakly-Supervised Video Anomaly Detection)<sup>2</sup> se utiliza una arquitectura transformer, como los videos tienen una dimensión temporal adicional para

observar todo el video, se necesita esta arquitectura transformer para luego dirigir la atención a cada porción de video para detectar anomalías. Primero el extractor de características toma el video original y lo segmenta creando un mapa de características para cada segmento, creando una amplificación de características FAM (Feature Amplification Module), donde calcula la norma, luego se integran características globales y locales sobre el transformador en lugar de simplemente maximizar la magnitud de las características anómalas y minimizar las normales, los resultados se observan en la figura 2.

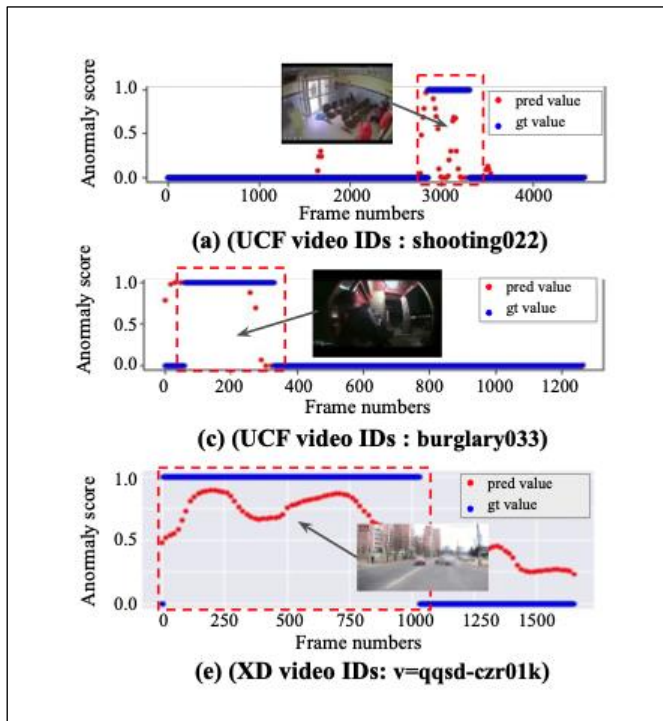


Figura 3 – Se observa que después de aplicar las transformaciones y la integración de características globales y locales (que son las predicciones de anomalías mostradas con puntos rojos en la figura), se tiene la predicción de las anomalías en el video.

## METODOLOGÍA

De acuerdo a la literatura y a algunos de los trabajos mencionados en este

reporte anteriormente relacionados a la clasificación de actividades anómalas en videovigilancia, este reporte se enfoca al artículo Learning Prompt-Enhanced Context Features for Weakly-Supervised Video Anomaly Detection<sup>5</sup>, ya que en una comparación con otros trabajos de la base de datos UCF\_crime para la detección de actividades anómalas, el trabajo citado en [5] es uno de los que obtuvieron alto valor de AUC, como a continuación se desarrolla, para esto se probaron los códigos proporcionados por el autor en el servidor del CIMAT, el Insurgente, ya que requieren un gran trabajo de cómputo y el uso de GPU para el procesamiento de los videos.

### Características de contexto mejoradas con prompts.

#### Módulos de Agregación del Contexto Temporal y aprendizaje mejorado con Prompts.

En la literatura se han definido diferentes métodos para la clasificación binaria de los videos que contienen una anomalía, los que están basados en aprendizajes por Múltiples Instancias (MIL), se les dificulta realizar una discriminabilidad más detallada cuando dentro de las anomalías hay una clasificación como en la base de datos UCF crime, para eso en Learning Prompt-Enhanced Context Features for Weakly-Supervised Video Anomaly Detection<sup>5</sup>, proponen un módulo de agregación de contexto temporal (TCA) y un módulo de aprendizaje mejorado por prompts PEL (Prompt-Enhanced Learning).

Primero divide el video en fragmentos no superpuestos mediante una ventana de 16 cuadros. Posteriormente aplica la red I3D<sup>6</sup> preentrenada que se basa en una red ConvNet 2D tradicional y se define como:

$$O(x, y) = \sum_{i=-k}^k \sum_{j=-k}^j W(i, j) * I(x + i, Y + j) \quad (2)$$

Donde:

- $W(i, j)$  es el valor del filtro en la posición  $(i, j)$ .
- $I(x+i, y+j)$  es el valor del píxel en la posición desplazada en la imagen de entrada.
- $O(x, y)$  es el valor del píxel en la posición  $(x, y)$  en la salida convolucionada.
- $k$  es el tamaño del filtro.

Y cuando se introduce la dimensión temporal la ecuación queda de la siguiente manera

$$O(x, y, t) = \sum_{i=-k}^k \sum_{j=-k}^k \sum_{m=-l}^l W(i, j, m) * V(x + i, y + j, t + m) \quad (3)$$

Donde:

- $W(i,j,m)$  es el valor del filtro 3D en la posición  $(i,j,m)$ , con  $m$  siendo la dimensión temporal.
- $V(x+i,y+j,t+m)$  es el valor del píxel en la posición desplazada en el video de entrada  $V$ .
- $O(x,y,t)$  es el valor del voxel (píxel volumétrico) en la posición  $(x,y,t)$  en la salida convolucionada.
- $l$  es el tamaño del filtro en la dimensión temporal.

Posteriormente aplica max Pooling 3D e implementa una arquitectura de flujo RGB y otra de flujo de flujo óptico representando la dinámica del video. Ambos flujos se procesan por separado mediante convoluciones 3D y sus salidas se fusionan para hacer una predicción conjunta expresado de la siguiente manera

$$F_{final} = F_{RGB} + F_{FLOW} \quad (4)$$

Y la salida final es:

$$\hat{y} = softmax(W + F_{final} + b) \quad (5)$$

Donde:

- $W$  son los pesos de la capa fully connected.
- $b$  es el sesgo.
- softmax convierte la salida en una probabilidad distribuida entre las clases de acción

El módulo de Agregación de contexto temporal (TCA), modela simultáneamente las dependencias globales y locales aprovechando una matriz de similitud y una sola fusión adaptativa.

Se obtiene una característica de contexto global aplicando tres capas lineales, dos  $f_q$  y  $f_k$  para obtener la matriz de similitud  $M$ , aplicándole softmax se obtiene el mapa de atención global y por último se le aplica la tercera capa lineal  $f_v$  para obtener la característica de contexto global  $X^g$ .

$$M = f_q(X) * f_k(X)^T \quad (6)$$

$$A^g = softmax\left(\frac{M}{\sqrt{D_h}}\right) \quad (7)$$

$$X^g = A^g * f_v(X) \quad (8)$$

Así mismo se determinan las características de calibración local que consideran el tiempo de ejecución del video.

$$\widetilde{M}_{IJ} = M_{IJ}, \text{ if } j \in \left[ \max(0, i - \left\lfloor \frac{\omega}{2} \right\rfloor), \min(i + \left\lfloor \frac{\omega}{2} \right\rfloor, T) \right] \quad (9)$$

$-\alpha,$       *para todo lo demas*

$$A^l = \text{softmax} \left( \frac{\widetilde{M}}{\sqrt{D_h}} \right) \quad (10)$$

$$X^l = A^l * f_v(X)$$

Donde  $w$  es el tamaño de la ventana de enmascaramiento y  $T$  es la longitud máxima de la secuencia de entrada, lo que permite capturar de manera efectiva pequeños cambios y lograr una mejora de características en el vecindario local.

Se implementa un perceptrón multicapa (MLP) de dos capas para reducir las características y obtener representaciones semánticas de alto nivel. Cada capa esta seguida por una activación GELU y una operación de dropout. Luego se utiliza una capa de convolución causal para predecir las puntuaciones de animalia a nivel fragmento.

El Aprendizaje Mejorado con Prompts (PEL) es un enfoque que enriquece las representaciones visuales incorporando información contextual basada en conocimientos, mejorando así la detección de anomalías en escenarios complejos.

Se lleva a cabo en tres partes, primero se realiza la construcción de prompts, se seleccionan relaciones comunes de ConcepNet como: IsA - Es un/una, PartOf - Parte de, HasA - Tiene un/una, UsedFor - Se usa para, CapableOf - Es capaz de, AtLocation - Se encuentra en, Causes - Causa, HasProperty - Tiene la propiedad de, Antonym - Antónimo de, Synonym - Sinónimo de, RelatedTo - Relacionado con, CausesDesire - Causa el deseo de, para construir un diccionario de conceptos manteniendo las de mayor relevancia, que sirven como propts para enriquecer las características visuales del modelo. Posteriormente se realiza la extracción de representaciones con CLIP un modelo multimodal que relaciona imágenes y texto mapea las imágenes y el texto a un espacio de características compartido, permitiendo al modelo utilizar la información textual para mejorar las predicciones visuales. Y al final las representaciones de los prompts generadas por CLIP se alinean con las

características visuales mejorando la discriminación de anomalías sin necesidad de emparejar bidireccionalmente texto e imagen durante la prueba. Se emparejan las características del primer plano de videos anómalos con prompts anómalos correspondientes, y las características de fondo con prompts normales.

## **Pruebas realizadas en el servidor del CIMAT, El Insurgente.**

Para las pruebas se necesita el uso de GPU, lo cual se solicitó el uso de los servidores de CIMAT otorgándonos acceso al del Insurgente, y Puerto Interior.

El proceso de entrenamiento original “Learning Prompt-Enhanced Context Features for Weakly-Supervised Video Anomaly Detection”<sup>5</sup> se completó en 117 minutos y 14 segundos, alcanzando un Área Bajo la Curva (AUC) de 0.8491, lo que indica una buena capacidad del modelo para distinguir entre clases positivas y negativas. Además, se logró una Tasa de Falsos Positivos (FAR) de 0.00264, lo que sugiere que el modelo tiene una baja probabilidad de generar falsas alarmas, haciendo que sea confiable para su aplicación en escenarios de detección de anomalías en videovigilancia.

Esta descripción resalta tanto la eficiencia del proceso de entrenamiento como la efectividad del modelo entrenado en términos de su capacidad de discriminación y su precisión en la identificación de eventos anómalos.

## **Implementación de Aumento de Datos**

El concepto de **Aumento de datos** es una técnica ampliamente utilizada en el campo del aprendizaje automático y, en particular, en la visión por computadora. Su propósito principal es mejorar la capacidad de generalización de los modelos aumentando la diversidad del conjunto de datos sin necesidad de recopilar nuevos datos. Esto se logra aplicando transformaciones a los datos existentes, creando versiones modificadas que aún representan la misma clase o categoría.



### Jittering Temporal (Temporal Jittering)

Se refiere a desplazar la secuencia de fotogramas hacia adelante o hacia atrás en el tiempo de manera aleatoria. Este desplazamiento temporal simula variaciones en el inicio o fin de una acción, lo que ayuda al modelo a ser más robusto frente a diferentes fases temporales de la acción.

### Corte Temporal (Temporal Cropping):

Consiste en seleccionar una sub-secuencia más corta de la secuencia original, simulando la situación en la que solo una parte de la acción está disponible o visible.

Al implementar data augmentation en el trabajo de “Learning Prompt-Enhanced Context Features for Weakly-Supervised Video Anomaly Detection”<sup>5</sup> se obtuvieron los resultados de la siguiente tabla 1, comparándolos con los resultados iniciales sin data augmentation.

Aspecto	Primer Entrenamiento	Segundo Entrenamiento	Comentarios
<b>Duración del Entrenamiento</b>	117 minutos y 14 segundos	344 minutos y 49 segundos	El segundo entrenamiento tomó significativamente más tiempo, probablemente debido al uso de Data Augmentation y operaciones adicionales.
<b>AUC (Área Bajo la Curva ROC)</b>	0.8422	0.8491	El AUC mejoró ligeramente en el segundo entrenamiento, indicando una mejor capacidad de generalización del modelo.
<b>Tasa de Falsos Positivos (FAR)</b>	0.00309	0.00264	La FAR fue menor en el segundo entrenamiento, lo que sugiere que el modelo cometió menos errores de clasificación.
<b>Impacto del Data Augmentation</b>	No se aplicó	Se aplicó (Temporal Jittering y Cropping)	Las técnicas de Data Augmentation en el segundo entrenamiento mejoraron tanto el AUC como la FAR.

*Tabla 1 : Resultados entre el código original y aplicando data Augmentation.*

### Trabajo a futuro.

Aún se está trabajando en la mejora y optimización de los códigos para la clasificación de actividades anómalas en videovigilancia para la seguridad ciudadana, en la siguiente tabla 2 se muestran algunas propuestas de mejorar para el código probado y otros más que ya están en la literatura.

Cambio Sugerido	Beneficio Esperado	Nivel de Impacto	Dificultad de Implementación
<b>Uso de Self-Attention Más Avanzado</b>	Capturará mejor las relaciones a largo plazo en las secuencias, crucial para el reconocimiento de acciones en videos.	Muy Alto	Alta
<b>Implementar Regularización (Dropout)</b>	Reducirá el sobreajuste, mejorando la generalización en datos no vistos.	Medio	Baja
<b>Experimentar con Modelos Preentrenados</b>	Aprovechar modelos preentrenados puede mejorar significativamente la precisión sin necesidad de entrenar desde cero.	Muy Alto	Alta
<b>Optimización del Proceso de Entrenamiento</b>	Ajustar el learning rate, usar estrategias de warm-up o learning rate scheduler puede acelerar la convergencia y mejorar la precisión.	Medio	Media
<b>Combinación de Características Multimodales</b>	Al combinar características de distintas fuentes, se capturan más aspectos del problema, lo que puede mejorar la precisión.	Alto	Alta

*Tabla 2: Mejoras propuestas como trabajo a futuro.*

## CONCLUSIONES

La clasificación de actividades anómalas en sistemas de videovigilancia requiere técnicas avanzadas que demandan un procesamiento intensivo de video, por lo que fue fundamental el acceso de los servidores de CIMAT. Es fundamental abordar de manera separada la extracción de características de los videos y utilizar estas características como entradas para la clasificación de comportamientos normales y anómalos. Además, dentro de las características anómalas, es necesario identificar y clasificar diferentes tipos de eventos, como robos u otras actividades ilícitas mencionadas en la base de datos UCF-Crime. Estos resultados pueden ser aplicados a videos obtenidos en entornos urbanos para mejorar la seguridad ciudadana y facilitar la respuesta ante situaciones de riesgo.

## REFERENCIAS

- [1] Lv, H., Zhou, C., Cui, Z., Xu, C., Li, Y., & Yang, J. Localizing anomalies from weakly-labeled videos. *IEEE transactions on image processing*, 30, 4505-4515, (2021)
- [2] Sultani, W., Chen, C., & Shah, M. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6479-6488). (2018).
- [3] Zhong, J. X., Li, N., Kong, W., Liu, S., Li, T. H., & Li, G. Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1237-1246). (2019).
- [4] Chen, Y., Liu, Z., Zhang, B., Fok, W., Qi, X., & Wu, Y. C. Mgfn: Magnitude-contrastive glance-and-focus network for weakly-supervised video anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 37, No. 1, pp. 387-395). (2023, June).
- [5] Pu, Y., Wu, X., & Wang, S. Learning Prompt-Enhanced Context Features for Weakly-Supervised Video Anomaly Detection. *arXiv preprint arXiv:2306.14451*, (2023).
- [6] Carreira, J., & Zisserman, A. Quo vadis, action recognition? a new model and the kinetics dataset. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6299-6308). (2017).
- [7] Simonyan, K., & Zisserman, A. *Two-stream convolutional networks for action recognition in videos*. In *Advances in neural information processing systems* (pp. 568-576), (2014).

- [8] Sultani, W., Chen, C., & Shah, M. *Real-world anomaly detection in surveillance videos*. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6479-6488), (2018).
- [9] Gong, D., Liu, L., Le, V., Saha, B., Mansour, M. R., Venkatesh, S., & Hengel, A. V. D. *Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection*. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1705-1714). (2019).

# IMPACTO DE LAS INTERACCIONES HIDRODINÁMICAS Y ELECTROSTÁTICAS EN LA DINÁMICA DE CONTAGIO DE VIRUS. UNA PERSPECTIVA MESOSCÓPICA.

**K. Gonzales, S. Ruano, I.A. Segoviano y F. Alarcon**

*División de Ciencias e Ingenierías, Universidad de Guanajuato Lomas del Bosque 103, 37150 León, Gto. México e-mail: paco@fisica.ugto.mx webpage: <http://www.dci.ugto.mx/dif/index.php/profesores/mecanicaestadistica/drfrancisco>*

**Palabras clave:** Simulación Mesoscópica, Dinámica de Fluidos, Virus, Vesículas, Lípidos, Proteínas, Adsorción, Difusión.

**Resumen.** En este documento reportamos los resultados obtenidos de las simulaciones mesoscópicas realizadas sobre la infraestructura de supercómputo de los laboratorios del Centro de Investigación en Matemáticas (CIMAT), gracias a dichas simulaciones logramos entender distintos aspectos de la dinámica y de la estructura de virus esféricos como el coronavirus, así como la dinámica de infección sobre la superficie de tejidos infectantes de modelos de fagocitos. Dada la naturaleza de las simulaciones mesoscópicas y al acceso a esta infraestructura fuimos capaces de encontrar tres resultados importantes en tres líneas del proyecto sometido. Por un lado, fuimos capaces de encontrar los parámetros mínimos necesarios para la formación de vesículas similares a la membrana de los coronavirus, llegando a simular las ecuaciones de movimiento de hasta 2 millones de partículas por simulación. Pudimos estudiar el efecto de campos hidrodinámicos oscilatorios en la difusión rotacional de coronavirus modelo en función del número de proteínas pico y la dinámica de adsorción de fagocitos en modelos de membranas celulares tomando en cuenta las proteínas de membranas, interacción fundamental para entender la dinámica de contagio virus-célula.

## 1. INTRODUCCIÓN

Ante la coyuntura y los cambios de paradigma que han abierto situaciones como el COVID-19 resulta más que pertinente este proyecto, el cual conjunta la experiencia de los investigadores y estudiantes involucrados en la aplicación de metodologías mesoscópicas a sistemas víricos modelo. Con la problemática actual de entender la dinámica de contagio de algunos tipos de virus. Estos resultados podrán permitir el diseño dispositivos tecnológicos que sirvan para la detección oportuna de posibles esparcimientos de virus o el diseño de nuevos fármacos que eviten pandemias como la del 2020.

A medida de que el virus SARS-CoV2 se propagó por viajeros y se volvió una pandemia que obligó el distanciamiento social debido a la suposición de que la principal ruta de infección era la transmisión de persona a persona por medio de las gotas de saliva, obligando el uso de cubrebocas y que debido a que las ciudades más pobladas presentaban los mayores números de infección. Para probar dicha hipótesis, empezaron a emerger investigaciones al respecto, por ejemplo Calles y Moran publicaron un artículo donde presentan un modelo para la carga viral expedida durante eventos respiratorios.<sup>1</sup> Ellos calcularon el número de viriones en gotas de saliva, que tan lejos viajan y cual es su dinámica en el aire, finalmente determinaron la dinámica de las partículas asumiendo que la gravedad y el arrastre son las principales fuerzas que dirigen el movimiento. Mediante la resolución numérica de las ecuaciones de movimiento que describen la dinámica de cada rango analizado, encontraron que en el régimen de aerosol las fuerzas de fricción reducen el movimiento horizontal de las gotas y caen extremadamente lento tiradas por la fuerza gravitacional en un tiempo de 3.33 min hasta incluso 3.83 días, dependiendo del flujo de aire. Entonces queda claro que para entender las gotas de virus y en general cualquier suspensión de virus se vuelve una tarea de altísima importancia resolver la dinámica de fluidos de dichos sistemas, es conocido que resolver dichas ecuaciones puede ser complicado incluso de manera numérica, en éste proyecto queda de manifiesto que una alternativa es mediante una perspectiva mesoscópica utilizando metodologías de resolución de dinámica de fluidos basada en partículas *microscópicas* de grano grueso, siempre y cuando se tenga acceso a recursos computacionales competitivos.

En este reporte se informa sobre los resultados de simulaciones

mesoscópicas asociados a tres aspectos importantes de suspensiones víricas, **en primer lugar** la simulación de suspensiones de moléculas lipídicas anfífilas para estudiar sistemáticamente el autoensamblado de vesículas, dichas vesículas se usan en ingeniería como réplicas de una o más funciones de células biológicas y en este caso también cierto tipo de virus cuya envoltura es una membrana bilipídica como en el caso del coronavirus.<sup>2,3</sup> **En segundo lugar** estudiamos la respuesta rotacional de un coronavirus modelo a un flujo oscilante en función de su número de proteínas pico (*Spike Proteins*) también conocidos como *peplómeros*,<sup>4</sup> este estudio es muy importante dado que la no motilidad de los virus hace que su movimiento se derive esencialmente de su propio movimiento térmico aleatorio y de los campos hidrodinámicos del medio en el que se encuentran embebidos, estos dos efectos modifican el movimiento rotacional del virus haciendo alinear sus peplómeros con las proteínas blanco de la célula a infectar. Entonces se puede decir que las propiedades que gobiernan el transporte y el éxito de infección de un virus es la difusión rotacional del mismo. Muy poca difusión rotacional en el virus provocará pocos alineamientos para infectar exitosamente, mientras que mucha difusión rotacional hará que los intervalos de alineamiento sean muy cortos para infectar con éxito.<sup>5</sup> **En tercer lugar** también estudiamos la adsorción de otro tipo de virus, los bacteriofagos, dichos virus típicamente infectan a las bacterias y a algunos protozoarios. Estos virus no son esféricos sino tienen una forma muy particular, los fagos que estudiamos en este proyecto son fagos con *cola* los cuales poseen unas proteínas que permiten inyectar el material genético dentro de la célula huésped.<sup>6</sup> Estos virus tienen la particularidad de infectar células motiles, como las bacterias con flagelos, así que la dinámica de contagio de dichos virus se ve influenciada por el campo hidrodinámico generado por los flagelos de las bacterias, la mucosa y las proteínas blanco, la temperatura, el pH, entre otros.

Este reporte está organizado de la siguiente manera. En la sección 2 presentamos que son los métodos y el modelado mesoscópico, posteriormente detallamos la metodología mesoscópica DPD utilizada en los tres sistemas simulados en éste proyecto, la ventaja de su utilización de manera paralelizada en la infraestructura del CIMAT y finalmente los detalles de simulación de cada uno de los tres sistemas. En la sección 3 mostramos y analizamos los resultados obtenidos para cada sistema, los cuales están organizados por subsecciones para cada

caso. Finalmente, en la sección 4 discutimos nuestras conclusiones del trabajo y algunas perspectivas del mismo.

## 2. METODOLOGÍA Y DETALLES DE SIMULACIÓN

### 2.1. Modelado mesoscópico

El modelado mesoscópico se concentra en predecir las propiedades materiales que están determinadas por las estructuras y comportamiento entre las escalas microscópicas (atómicas) y macroscópicas (continuo). Esta escala, típicamente se encuentra entre 10 - 1000 nm y 1ns - 10ms, es de particular importancia para el entendimiento de sistemas poliméricos, soluciones con surfactantes, bio-moléculas y otros fluidos complejos. Un fluido complejo puede ser descrito como un fluido en el que el comportamiento observable es afectado de manera fundamental, por la estructura microscópica del fluido. La cual puede ser una estructura molecular no trivial, por ejemplo un polímero o la presencia de una suspensión, como los coloides. Tales sistemas son de gran interés práctico debido a su aplicación en áreas tan diversas como la exploración de petróleo, desarrollo de recubrimientos, pinturas, procesamiento de alimentos y en medicina. Presentan un problema particularmente difícil debido a la disparidad inherente entre las escalas de longitud y tiempo involucradas en el sistema. Fenómenos físicos ocurren en escalas temporales y espaciales macroscópicas, mientras que las causas fundamentales de estos fenómenos pueden ocurrir en la propia escala de los elementos que forman el fluido complejo.<sup>7</sup> Este rango de escalas tan amplio representa un obstáculo significativo para muchas técnicas de simulación tradicionales.

Es bien conocido que un fluido simple puede ser, bajo muchas condiciones experimentales, adecuadamente descrito a escala macroscópica por ecuaciones como las de Navier - Stokes.<sup>8</sup> Existen sofisticadas técnicas computacionales para la solución de estas ecuaciones diferenciales.<sup>9</sup> Sin embargo, estos métodos pueden incluir solamente los detalles microscópicos del fluido de una manera fenomenológica y son entonces, malos candidatos para aplicarlos a fluidos complejos.

En una escala microscópica, los métodos de dinámica molecular<sup>10</sup> permiten la simulación de líquidos mediante el seguimiento de la posición y el momento de cada partícula en el fluido. A pesar de que las aproximaciones de dinámicas moleculares se han ido incrementando



exitosamente en la simulación de un relativamente pequeño número de partículas,<sup>11</sup> intentos para aplicar estas técnicas microscópicas a sistemas de fluidos complejos muchas veces resultan en necesidades computacionales irreales.

Mientras que los métodos computacionales convencionales como dinámicas moleculares en escalas microscópicas y dinámicas de fluidos en escalas macroscópicas han sido conocidos por sus éxitos considerables, la escala mesoscópica de los fluidos complejos todavía se mantiene lejos de su alcance. Las razones se deben a los requerimientos de una gran cantidad de poder computacional para el caso de las dinámicas moleculares microscópicas y a la ruptura de la descripción del continuo para el caso de las dinámicas de fluidos macroscópicas. Esta situación ha disparado el desarrollo de métodos mesoscópicos específicamente destinados a llenar el hueco entre las aproximaciones micro y macroscópica.

En la última década del siglo XX se alcanzó un gran progreso en el área de simulación mesoscópica. Entre las técnicas sobresalientes se encuentran: el método de Lattice - Boltzmann (LBM),<sup>12</sup> dinámica del funcional de la densidad<sup>13,14</sup> y la familia de métodos basados en partículas de grano grueso como el caso de la dinámica de partícula disipativa (DPD).<sup>15,16</sup>

La simulación mesoscópica no solamente ha tenido un área de investigación interesante para la comunidad académica durante los últimos años, la industria también ha mostrado gran interés en ello e incluso, han sido los responsables de muchos logros. Por ejemplo, el método DPD fue concebido en Shell<sup>15</sup> y luego desarrollado en Unilever.<sup>16</sup> También ha sido aplicado en la industria mexicana en Comex.<sup>17</sup>

## 2.2. Dinámica de Partículas Disipativas (DPD)

En DPD el fluido está compuesto de un conjunto de partículas puntuales en el espacio continuo y el movimiento de las partículas está regido por las ecuaciones de Newton. Este es un marco común para todos los métodos de simulación basados en partículas, como las dinámicas moleculares (DM).<sup>10,11</sup> Lo que hace diferente a DPD es la interpretación física que tienen las partículas y la forma específica de las fuerzas entre partículas. La interpretación de las partículas en DPD es que no representan moléculas individuales o átomos dentro del sistema, sino cada una, describe la posición y el momento de una

región del fluido, el contorno exacto de cada partícula no está definido por el modelo. Se considera que ésta construcción de grano grueso del sistema permite una descripción de escalas de longitudes mesoscópicas. El sistema DPD consiste en un conjunto de  $N$  partículas, las cuales se encuentran en el espacio continuo. Cada partícula DPD está descrita por su vector de posición  $\mathbf{r}_i$  y su momento  $\mathbf{p}_i$ , juntos con su masa  $m_i$ , donde  $i$  será usada para distinguir entre las partículas. Seguiremos la convención de suponer que todas las partículas tienen la misma masa  $m_i = m$ , usaremos también  $\mathbf{v}_i$  para representar a la velocidad de una partícula. Esta velocidad está definida a través de la definición normal del momento  $\mathbf{p}_i = m\mathbf{v}_i$ . La evolución del algoritmo para DPD fue originalmente descrito<sup>15</sup> en la forma de ecuaciones estocásticas de pasos de tiempo finito para la evolución de las posiciones y los momentos de cada partícula. Sin embargo, esto ha ido interpretado de manera más útil como una evolución infinitesimal temporal.<sup>16</sup>

### 2.3. LAMMPS y la simulación en paralelo

Ante los requerimientos de computación de alto rendimiento para los sistemas biológicos que queremos modelar, hemos tomado la decisión de usar un simulador masivamente paralelizado de simulación molecular con el cual hemos tenido experiencia trabajando desde el 2020 a la fecha en computadoras de alto rendimiento con éxito comprobado.<sup>18–23</sup> LAMMPS es el acrónimo de *Large-scale Atomic/Molecular Massively Parallel Simulator*, LAMMPS es un código de simulación de dinámica molecular clásica enfocado en modelamiento de materiales.<sup>24</sup> Fue diseñado para correr de manera eficiente en computadoras en paralelo y para ser fácilmente extendido y modificado, esto en gran medida al hecho de que es código-abierto distribuido bajo los términos del GNU Public License Version 2 (GPLv2). LAMMPS puede ser compilado y corrido sobre máquinas de escritorio o laptops personales y correr incluso en serie, pero está diseñado para computadoras paralelizadas de alto rendimiento que soporte las librerías de paso de mensaje MPI.<sup>25</sup> Esto incluye memoria compartida de multi-núcleo, servidores de multi-CPU, clusters con memoria distribuida y supercomputadoras. Algunas partes de LAMMPS inclusive soportan OpenMP de multi-hilos, vectorización y aceleración en GPU. Este software está escrito en C++ y requiera un compilador que sea al

menos compatible con el estandar C++-11.

Sobre maquinas en paralelo, LAMMPS usa técnicas de descomposición espacial con paralelización MPI para particionar la simulación en sub-simulaciones del mismo costo computacional, a los que cada uno es asignado a cada procesador. Los procesadores se comunican y guardan información de átomos “fantasmas” para los átomos que estan en la frontera entre las sub-simulaciones.

#### 2.4. Detalles de simulación vesículas

Para modelar la molécula fosfolipídica, enlazamos las partículas DPD de la molécula mediante un potencial armónico:

$$V(r_{ij}) = \frac{1}{2}K(r_{ij} - l_0)^2 \quad (1)$$

donde  $K$  es la constante de resorte y  $l_0$  es la longitud de enlace de equilibrio. Además, introducimos un enlace con rigidez en la flexión del fosfolípido. Para explorar los efectos de rigidez sobre las estructuras resultantes. Desarrollamos dos simulaciones por separada usando interacciones de rigidez de flexión diferentes. A los que llamamos  $B_A$  y  $B_B$  para diferenciarlos. De manera que,

$$B_A(\theta) = \frac{1}{2}k_\theta(\theta - \theta_0), \quad B_B(\theta) = k_\theta(1 - \cos(\theta - \theta_0)) \quad (2)$$

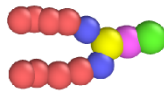


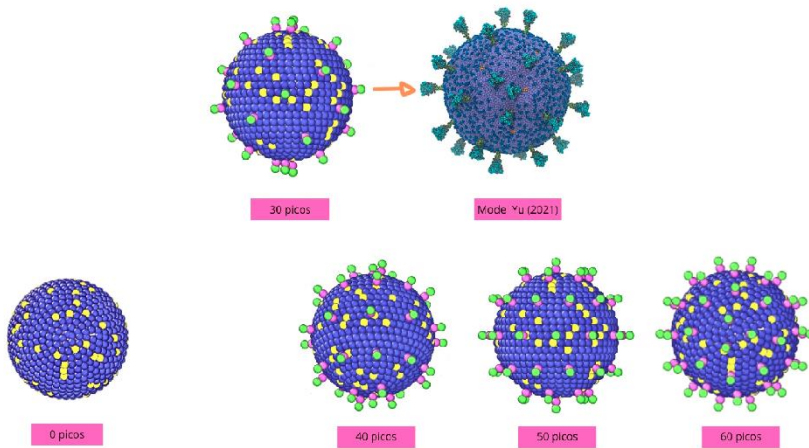
Figura 1: Mapeo de molécula de dimyristoyl phosphatidyl choline (DMPC).

Utilizamos un mapeo de la molécula lipídica real llamado *forma-Y invertida*, como la llama originalmente Kranenburg et al.,<sup>26</sup> a partir de la Figura 1 podemos distinguir las partículas del fosfolípido que representan la parte hidrófoba en color rojo y azul, mientras que la parte hidrófila en el resto de colores.

Sobre el detalle del mapeo, podemos distinguir que en rosa se representa la parte carbonatada, donde cada bola roja representa tres carbonos con sus respectivos hidrogenos, las bolas azules representan la parte carboxilada, mientras que la bola rosa y verde representan la parte del fósforo y el nitrógeno respectivamente y la amarilla los cuatro carbonos que sirven de puente entre los carboxilos y el residuo fosfato.

## 2.5. Detalles de simulación coronavirus

El modelo consiste en una suspensión diluida de modelos de virus, que están inmersos en un fluido DPD, este fluido es posicionado entre dos paredes paralelas y está sometido a una fuerza cortante oscilatoria.



*Figura 2: Esquema de los modelos de virus usados en cada simulación, donde el modelo referencial de 30 picos es tomado del trabajo de Yu et al., 2021.<sup>27</sup>*

En nuestro modelo, el virus tiene un radio de  $4r_c$ , donde  $r_c$  es la unidad de longitud en DPD<sup>16</sup> y contiene picos colocados de manera explícita en la superficie, esta cantidad de picos varía de manera aproximadamente simetría (mirar la Figura 2), por lo que, la respuesta del virus bajo distintas amplitudes del flujo cortante oscilatorio, se estudia para todos las concentraciones de picos.

El tamaño adimensional de la caja de simulación es de  $30 \times 21 \times 30$  y el número total de partículas DPD es  $N \geq 81\,370$ , donde la densidad numérica es  $\rho \geq 3$  a temperatura ambiente.

Para cada simulación se usó 10 nodos, 120 hilos y se corrió hasta  $1 \times 10^7$  pasos de tiempo. Se guardó los datos estadísticos desde los últimos  $8 \times 10^6$  pasos de tiempo para calcular el coeficiente de difusión rotacional de cada modelo rígido de virus.

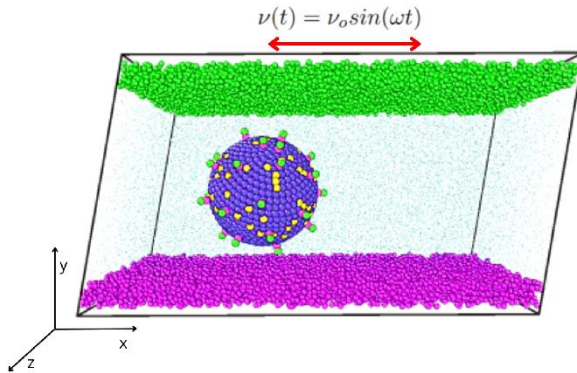


Figura 3: Esquema del sistema, donde el virus experimenta un flujo cortante oscilatorio entre dos paredes paralelas. Las flechas rojas indican la dirección de las velocidades de las paredes, que están sujetas a una deformación por fuerza cortante determinada por  $v(t)$ .

Como se mencionó anteriormente, el fluido DPD se encuentra entre dos paredes paralelas, éstas son hidrófilas y tienen una separación de  $17r_c$ , ambas se mueven a la misma velocidad pero en sentidos opuestos, ya que están sujetas a la deformación oscilatoria aplicada al sistema mediante el comando *fix move wiggle* de LAMMPS<sup>24</sup> (mirar la Figura 3).

## 2.6. Detalles de simulación fagos

Para la modelización del fago, aproximamos la cabeza del virus como una bola DPD (bola violeta en la Figura 4) con propiedades hidrófilas y la cola del virus como otra bola DPD (bola verde en la Figura 4), la anisotropía de la interacción del virus con la bacteria u

otro virus está dada por los valores de intensidad de la interacción de la fuerza conservativa de la dinámica de simulación. Adicionalmente, tenemos un modelo de enlace entre la cabeza y la cola de los virus como una fuerza armónica dada por

$$F_h(\vec{r}) = K(\vec{r}_h - \vec{r}_t), \quad (3)$$

donde tenemos que

$$l_0 = |\vec{r}_h - \vec{r}_t|^2, \quad (4)$$

y  $K$  es la magnitud de la fuerza de enlace, mientras que  $l_0$  es la longitud de equilibrio del enlace cabeza-cola del virus.

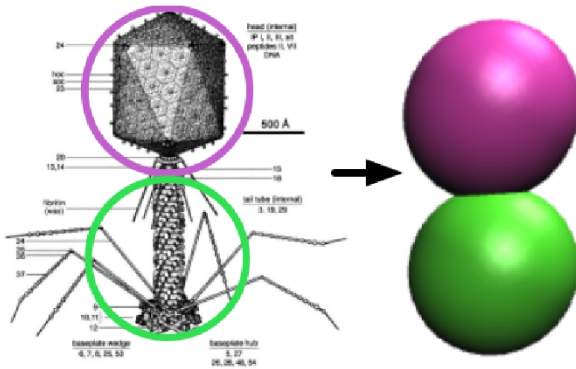
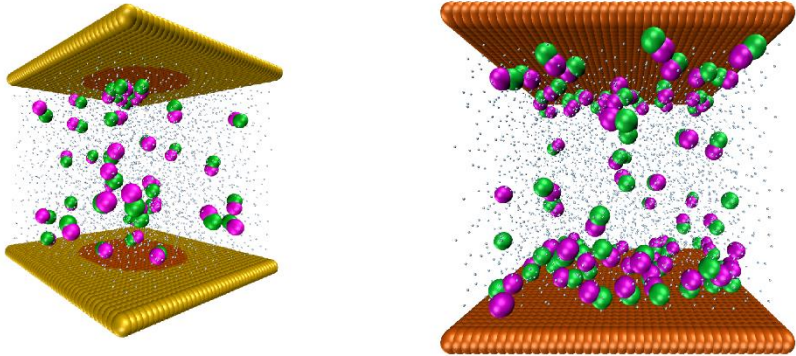


Figura 4: Bacteriofago modelado como un dímero, donde las bolas se conectan por medio de un enlace armónico con parámetro armónico  $K$ . Figura del virus tomada de.<sup>28</sup>

En las simulaciones construimos dos paredes con propiedades hidrofílicas que representarán la superficie de la membrana de la bacteria a infectar por los virus sobre el plano  $XZ$ . En la Figura 5 podemos observar que las paredes amarillas representan la parte de los fosfolípidos de la membrana, mientras que las bolas marrones en las paredes representan las proteínas blanco transmembranales que son más afines a las colas de los fagos. Las colas de los fagos están representadas por bolas verdes mientras que las cabezas por bolas rosas. Las bolas pequeñas son las aguas que están representadas así solamente para ayudar a la observación de los virus.

Con la ayuda de la infraestructura del CIMAT se lograron realizar estudios sistemáticos de suspensiones víricas a distintas cargas virales



*Figura 5: Fotos de dos simulaciones con dos concentraciones de proteína blanco diferentes.*

y concentraciones de proteínas blanco. En la sección de Resultados se muestra las mediciones de adsorción de virus en las paredes y como esta influye en la dinámica de contagio de la bacteria. Tomamos en cuenta la afinidad de las colas de los virus con las proteínas blanco mayor a la afinidad que puedan tener con la membrana celular, del mismo modo que la cabeza del virus tiene una menor afinidad con las mismas proteínas, básicamente por la naturaleza electrostática de dicha interacción entre proteínas.

### 3. RESULTADOS

#### 3.1. Autoensamblado de vesículas

En cuanto a las simulaciones, se realizaron simulaciones DPD en el ensamble NVT, donde se utilizaron 800 moléculas de fosfolípidos embebidos en solvente modelado explícitamente como bolas DPD. Estas simulaciones fueron llevadas a cabo a una temperatura reducida de  $T^* = 1$  dentro de una caja de simulación de tamaño  $L^3$  correspondiente, de tal manera que la densidad de bolas de DPD se mantuviera constante a  $\rho = 3$ , para que el comportamiento hidrodinámico se

mantuviera válido. Para cada simulación variamos sistemáticamente la concentración de moléculas de lípidos, tal que la concentración de definió como

$$\phi = \frac{N_{LP}}{N_T}, \quad (5)$$

donde  $N_T$  es el número total de bolas DPD y  $N_{LP}$  es el número de bolas de lípido que tenemos. Usando esta definición, determinamos el tamaño apropiado de la caja cúbica. Para mantener la densidad deseada de bolas se añadían las bolas de agua correspondientes.

A continuación se muestran los resultados de las simulaciones después de que el sistema se ha auto-ensamblado y llegado a un estado estacionario. Cada subpanel, tanto en la Figura 6 como en la Figura 7 tienen puesto el valor de  $\phi$  correspondiente

En la Figura 6 podemos notar la formación de una vesícula a una concentración  $\phi = 0,11$  y  $\phi = 0,09$ , sin embargo dicha estructura coexiste con tubos y con bicapas como lo podemos observar en paneles (g-d) de la misma figura. Otro resultado aparece a densidades más bajas donde ahora podemos observar la formación de dos vesículas de la mitad de tamaño a  $\phi = 0,07$ , sin embargo, al igual que a densidades intermedias, coexisten con tubos y bicapas. Incluso para densidades aún más bajas de 0,05 coexisten vesículas más pequeñas con la formación de tubos del tamaño de la caja. En el régimen de concentraciones intermedias  $0,1 < \phi < 0,2$ , podemos notar que el sistema tiende a la formación de tubos del tamaño del sistema, mientras que concentraciones aún más grandes los lípidos forman bicapas o incluso dos bicapas paralelas.

Para la Figura 7 podemos observar un comportamiento distinto, el cambio en la rigidez de flexión al modelo  $B_B N$  de las moléculas nos permite observar bicapas más flexibles a concentraciones altas en comparación con el modelo  $B_A$ . Para concentraciones intermedias la formación de tubos termina siendo más estable, mientras que para concentraciones bajas  $\phi \leq 0,1$  la formación de vesículas domina.

### 3.2. Difusión y rotación de modelos de virus esféricos

La aplicación de un flujo cortante oscilatorio sobre el sistema da como resultado una competencia entre la respuesta del fluido a la fuerza cortante oscilatoria y a la fuerza aleatoria, que mantiene el movimiento térmico de las partículas.



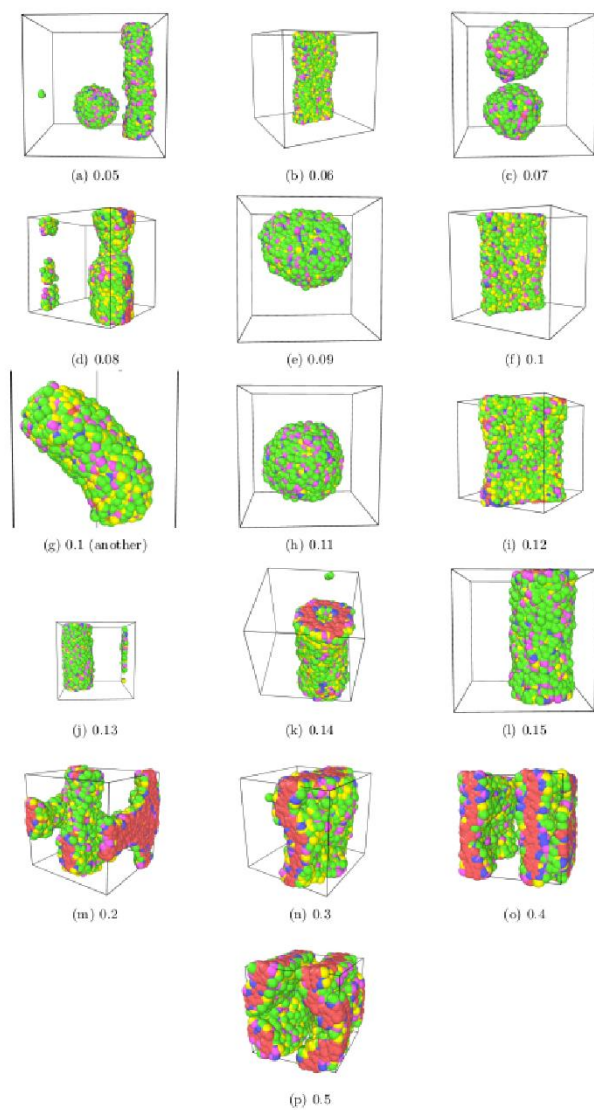
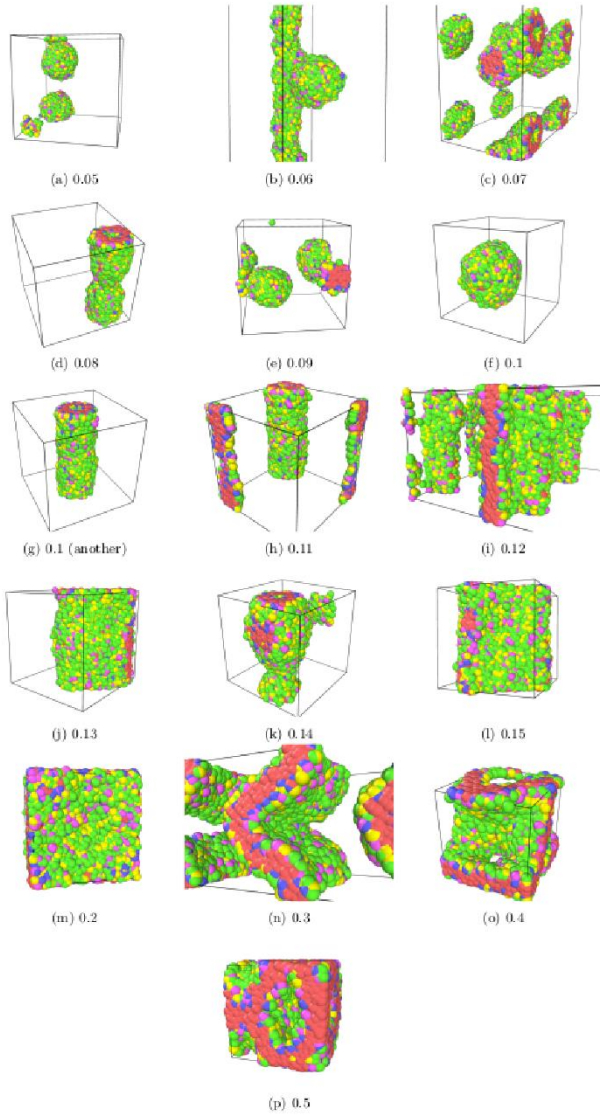


Figura 6: Diagrama de fase con rigidez de flexión  $B_A$  en términos de la concentración de lípidos.



*Figura 7: Diagrama de fase con rigidez de flexión  $B_B$  en términos de la concentración de lípidos*

La caracterización de nuestros resultados mediante el cálculo del

desplazamiento cuadrático medio rotacional ( $\langle \Delta\theta^2 \rangle^*$ ) nos brinda información sobre la difusión rotacional de los virus, ya que el objetivo es estudiar la respuesta de los virus cuando son sometidos a distintos tipos de regímenes debido a la presencia de un flujo cortante oscilatorio.

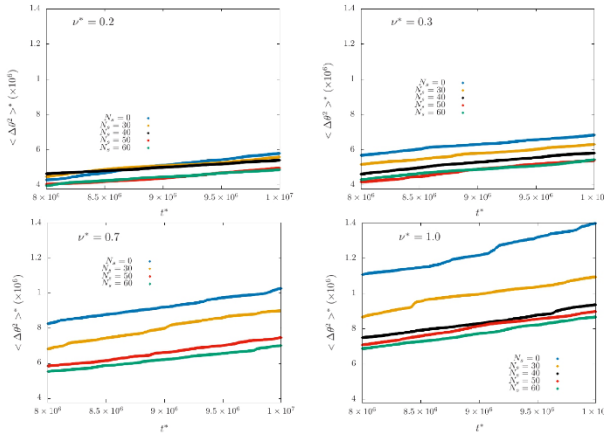


Figura 8: El desplazamiento cuadrático medio rotacional ( $\langle \Delta\theta^2 \rangle^*$ ) vs. el tiempo para 4 deformaciones de la fuerza cortante oscilatoria  $\nu^*$ . El número de picos es representado por  $N_s$ .

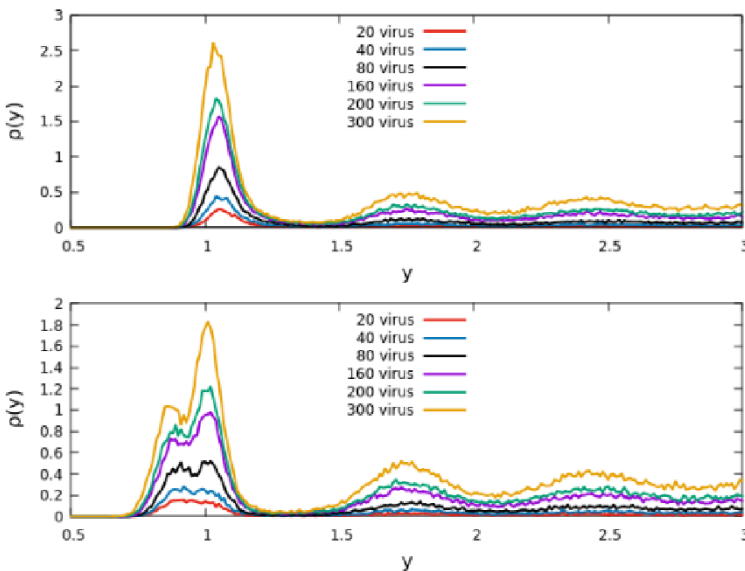
La Figura 8 muestra un aumento del desplazamiento cuadrático medio rotacional del virus cuando la deformación por el flujo cortante aumenta de  $\nu^* = 0.2$  a  $\nu^* = 1.0$ , en consecuencia, estas variables muestran una relación proporcional independiente del número de picos en la superficie del virus.

Por el contrario, en la misma Figura 8, el desplazamiento cuadrático medio rotacional disminuye con el aumento de la concentración de picos, este cambio es mayor cuando el flujo cortante domina sobre el movimiento térmico (ver las gráficas  $\nu^* = 0.7$  y  $\nu^* = 1.0$ ).

### 3.3. Adsorción de fagos

Primero que nada, note que los perfiles de densidad medidos en estas simulaciones muestran que las colas se adhieren más a la su-

perficie que las colas, tanto en el caso de  $\chi = 0,25$  como en el caso de  $\chi = 0,75$ . Sin embargo, cuando los virus están cerca de una area con poca proteína blanco, podemos notar que las colas de los virus forman dos zonas de concentración alta (Figura 9), observandose dos picos cercanos a la pared, el pico más cercano a la pared representa los virus adheridos a la proteína pero el segundo pico que aparece cuando la carga viral sobrepasa los 40 virus es debido a la acumulación de virus sobre la zona de la membrana que no es proteína, la cual comparte con las cabezas de esos mismos virus y de las cabezas de los virus adheridos a las proteínas, note que el segundo pico de las colas está a la misma distancia que el primer pico de las cabezas de los virus.



*Figura 9: Perfiles de densidad de partículas adheridas a la superficie de la célula, cuando se tienen una concentración de proteína del 25%. El panel superior corresponde al perfil de densidad de las cabezas de los virus en función de la distancia a la pared. El panel inferior corresponde al perfil de densidad de las colas de los virus en función de la distancia a la pared.*

Es interesante notar que en el caso de una zona con mayor cantidad de proteínas blanco (Figura 10) el mecanismo de adsorción se vea modificado, en este caso no existe la aparición de este segundo pico, en gran medida por el hecho de que los virus tienen más espacios accesibles a la adsorción. Otra consecuencia de que haya mayor adsorción será la falta de difusión de los fagos una vez lleguen a la proteína, por lo que se infiere que los virus quedan completamente adheridos a la superficie.

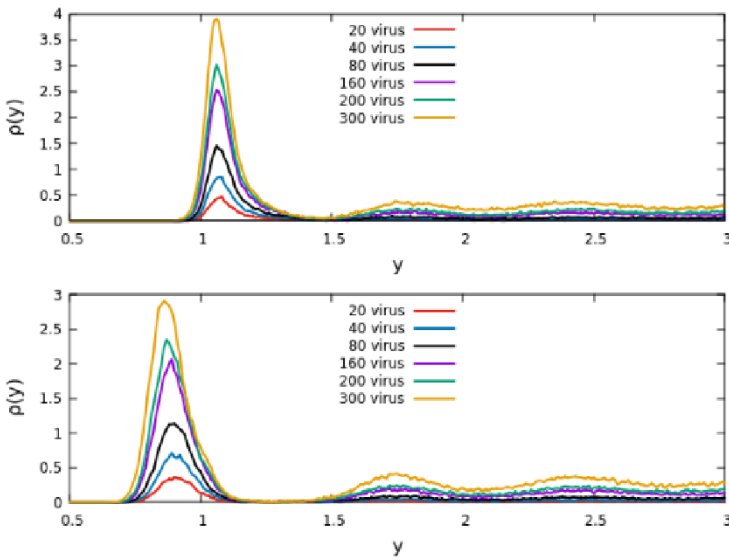


Figura 10: Perfiles de densidad de partículas adheridas a la superficie de la célula, cuando se tienen una concentración de proteína del 75%. El panel superior corresponde al perfil de densidad de las cabezas de los virus en función de la distancia a la pared. El panel inferior corresponde al perfil de densidad de las colas de los virus en función de la distancia a la pared.

## 4. CONCLUSIONES

A través de la utilización de simulaciones de Dinámica de Partícula Disipativa (DPD), investigamos distintas propiedades relacionadas con el autoensamblaje de vesículas de lípidos. Además, analizamos cómo la variación de los parámetros del DPD influye en este proceso. El objetivo principal fue comprender cómo los parámetros del DPD afectan la formación y estabilidad de las vesículas de fosfolípidos. De esta forma, los resultados obtenidos brindan información valiosa sobre los mecanismos de autoensamblaje y las propiedades de estas estructuras lipídicas.

Para el caso de coronavirus modelo, los resultados muestran que la difusión rotacional del virus es afectada por la población de picos en la superficie que cambia la propiedad estructural del modelo. Dicha rotación se ve afectada por la competición de dos fuerzas la térmica y la cortante oscilante cuando la amplitud del esfuerzo cortante es baja, por lo contrario, cuando la fuerza cortante oscilante domina, la rotación del virus es alta y sigue el flujo generado por la fuerza cortante.

Los bacteriofagos resultan un caso muy especial, donde los resultados de las simulaciones nos han dado luz sobre como los fagos infectan a las bacterias, sin embargo hace falta más analisis sobre la dinámica de los fagos una vez se acercan y se adhieren a la superficie de la celula, por ejemplo cómo es su difusión sobre la membrana o si la orientación de los fagos está relacionada con una mayor probabilidad de infección.

Finalmente comentar que las tres líneas del proyecto estan en proceso de escritura de un manuscrito cada uno para una proxima publicación. Comentar también que se pretende continuar con una solicitud de acceso al Supercomputo del CIMAT para el proximo año para investigar las preguntas pendientes sobre las interacciones electrostaticas en el caso del coronavirus y de los flujos hidrodinámicos en el caso de los fagos.

## REFERENCIAS

- [1] A Calles and J L Morán-López. Modeling the viral load expelled in saliva droplets carrying sars-cov-2. *Journal of Physics Com-*

- munications*, 5(9):095005, sep 2021.
- [2] Zhikuan Zhang, Norimichi Nomura, Yukiko Muramoto, Toru Ekimoto, Tomoko Uemura, Kehong Liu, Moeko Yui, Nozomu Kono, Junken Aoki, Mitsunori Ikeguchi, Takeshi Noda, So Iwata, Umeharu Ohto, and Toshiyuki Shimizu. Structure of sars-cov-2 membrane protein essential for virus assembly. *Nature Communications*, 13(1), 8 2022.
  - [3] Yixin Huang, Tongyun Wang, Lijie Zhong, Wenxin Zhang, Yu Zhang, Xiulian Yu, Shuofeng Yuan, and Tao Ni. Molecular architecture of coronavirus double-membrane vesicle pore complex. *Nature*, 8 2024.
  - [4] FRANK FENNER, PETER A. BACHMANN, E. PAUL J. GIBBS, FREDERICK A. MURPHY, MICHAEL J. STUDERT, and DAVID O. WHITE. Structure and composition of viruses. In *Veterinary Virology*, pages 3–19. Elsevier, 1987.
  - [5] M. A. Kanso, J. H. Piette, J. A. Hanna, and A. J. Giacomini. Coronavirus rotational diffusivity. *Physics of Fluids*, 32(11):113101, 11 2020.
  - [6] H.-W. Ackermann and H. M. Krisch. A catalogue of t4-type bacteriophages. *Archives of Virology*, 142(12):2329–2345, 12 1997.
  - [7] Masao Doi and Sam F Edwards. *The theory of polymer dynamics*. International series of monographs on physics. Oxford Univ. Press, Oxford, 1986.
  - [8] G. K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge University Press, 1967.
  - [9] Dale A. Anderson, John C. Tannehill, Richard H. Pletcher, Munipalli Ramakanth, and Vijaya Shankar. *Computational Fluid Mechanics and Heat Transfer*. CRC Press, 12 2020.
  - [10] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, USA, 5 2017.
  - [11] Daan Frenkel and B. Smit. *Understanding Molecular Simulation*. Academic Press, 11 2001.
  - [12] Shiyi Chen and Gary D. Doolen. Lattice boltzmann method for fluid flows. *Annual Review of Fluid Mechanics*, 30(Volume 30, 1998):329–364, 1998.
  - [13] J. G. E. M. Fraaije, B. A. C. van Vlimmeren, N. M. Maurits, M. Postma, O. A. Evers, C. Hoffmann, P. Altevogt, and G. Goldbeck-Wood. The dynamic mean-field density functional method and its application to the mesoscopic dynamics of quen-

- ched block copolymer melts. *The Journal of Chemical Physics*, 106(10):4260–4269, 03 1997.
- [14] T. KAWAKATSU, M. DOI, and R. HASEGAWA. Dynamic density functional approach to phase separation dynamics of polymer systems. *International Journal of Modern Physics C*, 10(08):1531–1540, 1999.
- [15] P. J. Hoogerbrugge and J. M. V. A. Koelman. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhysics Letters*, 19(3):155, jun 1992.
- [16] P. Español and P. Warren. Statistical mechanics of dissipative particle dynamics. *Europhysics Letters*, 30(4):191, may 1995.
- [17] A. Gama Goicochea, E. Nahmad-Achar, and E. Pérez. Colloidal stability dependence on polymer adsorption through disjoining pressure isotherms. *Langmuir*, 25(6):3529–3537, 2 2009.
- [18] Diego Rogel Rodriguez, Francisco Alarcon, Raul Martinez, Jorge Ramírez, and Chantal Valeriani. Phase behaviour and dynamical features of a two-dimensional binary mixture of active/passive spherical particles. *Soft Matter*, 16:1162–1169, 2020.
- [19] Josué Jara, Francisco Alarcón, Ajay K. Monnappa, José Ignacio Santos, Valentino Bianco, Pin Nie, Massimo Pica Ciamarra, Ángeles Canales, Luis Dinis, Iván López-Montero, Chantal Valeriani, and Belén Orgaz. Self-adaptation of pseudomonas fluorescens biofilms to hydrodynamic stress. *Frontiers in Microbiology*, 11, 2021.
- [20] José Martín-Roca, Raul Martinez, Lachlan C. Alexander, Angel Luis Diez, Dirk G. A. L. Aarts, Francisco Alarcon, Jorge Ramírez, and Chantal Valeriani. Characterization of MIPS in a suspension of repulsive active Brownian particles through dynamical features. *The Journal of Chemical Physics*, 154(16):164901, 04 2021.
- [21] Pin Nie, Francisco Alarcon, Iván López-Montero, Belén Orgaz, Chantal Valeriani, and Massimo Pica Ciamarra. In-silico modeling of early-stage biofilm formation. *Soft Materials*, 19(3):346–358, 3 2021.
- [22] Enrique Chacón, Francisco Alarcón, Jorge Ramírez, Pedro Tarazona, and Chantal Valeriani. Intrinsic structure perspective for mips interfaces in two-dimensional systems of active brownian particles. *Soft Matter*, 18:2646–2653, 2022.
- [23] José Martín-Roca, Valentino Bianco, Francisco Alarcón, Ajay K.



- Monnappa, Paolo Natale, Francisco Monroy, Belen Orgaz, Ivan López-Montero, and Chantal Valeriani. Rheology of *Pseudomonas fluorescens* biofilms: From experiments to predictive DPD mesoscopic modeling. *The Journal of Chemical Physics*, 158(7):074902, 02 2023.
- [24] Aidan P. Thompson, H. Metin Aktulga, Richard Berger, Dan S. Bolintineanu, W. Michael Brown, Paul S. Crozier, Pieter J. in 't Veld, Axel Kohlmeyer, Stan G. Moore, Trung Dac Nguyen, Ray Shan, Mark J. Stevens, Julien Tranchida, Christian Trott, and Steven J. Plimpton. Lammmps - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, 271:108171, 2022.
- [25] Sayantan Sur, Matthew J. Koop, and Dhabaleswar K. Panda. High-performance and scalable mpi over infiniband with reduced memory usage: an in-depth performance analysis. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, SC '06, page 105–es, New York, NY, USA, 2006. Association for Computing Machinery.
- [26] Marieke Kranenburg, Jean-Pierre Nicolas, and Berend Smit. Comparison of mesoscopic phospholipid–water models. *Phys. Chem. Chem. Phys.*, 6:4142–4151, 2004.
- [27] Alvin Yu, Alexander J. Pak, Peng He, Viviana Monje-Galvan, Lorenzo Casalino, Zied Gaieb, Abigail C. Dommer, Rommie E. Amaro, and Gregory A. Voth. A multiscale coarse-grained model of the sars-cov-2 virion. *Biophysical Journal*, 120(6):1097–1104, 3 2021.
- [28] P. G. Leiman, S. Kanamaru, V. V. Mesyanzhinov, F. Arisaka, and M. G. Rossmann. Structure and morphogenesis of bacteriophage t4. *Cellular and Molecular Life Sciences (CMLS)*, 60(11):2356–2370, 11 2003.