

ISSN 2007-5448



RECIBE

Revista electrónica

DE COMPUTACIÓN, INFORMÁTICA, BIOMÉDICA Y ELECTRÓNICA

Índice

Computación e Informática

- **Una Introducción amable pero riguroza al aprendizaje por refuerzo.** **C1**
Mauro Alejandro Montenegro Meza, Rolando Menchaca Méndez, Ricardo Menchaca Méndez
- **Hacia una propuesta de rúbrica para la evaluación de la calidad de comentarios de código en Java.** **C2**
Juan Carlos García Murillo
- **Prueba de usabilidad al módulo del terapeuta en un ambiente virtual para la rehabilitación motriz pos-ictus del miembro superior.** **C3**
Guadalupe Toledo Toledo, Aguilar Acevedo Francisco, Nieva García Omar Santiago, Daniel Pacheco Bautista, Bezares Molina Francisco Gabriel, Marleydy Acevedo Gómez
- **La presencia docente como determinante de la presencia cognitiva en un curso Blended Learning basado en la Comunidad de Indagación.** **C4**
Julio Erick Medina López, Héctor Martínez Ruíz, Ezra Federico Parra González, Jezreel Mejía Miranda
- **Chatgpt: La inteligencia artificial como herramienta de apoyo al desarrollo de las competencias STEM en los procesos de aprendizaje de los estudiantes.** **C5**
Yessica Rodríguez Almazán, Ezra Federico Parra-González, Kathia Anahí Zurita-Aguilar, Jezreel Mejía Miranda, David Bonilla Carranza
- **Revisión de Problemas en la Detección de Objetos en Imágenes y Videos Digitales.** **C6**
Miguel Ángel Gutiérrez Velázquez, Mario Chacón, Alma Delia Corral Sáenz
- **Sistema de Alerta Temprana epidemiológica animal.** **C7**
Edgar Montiel Cruz, Felipe Hernández González, Alejandro Luis Collantes Chávez Costa
- **El Control Neuro Proporcional-Derivativo para la formación Líder-Seguidor entrenado con el Filtro de Kalman Extendido.** **C8**
José de Jesús Hernández Barragán, Alma Yolanda Alanís García, Erasmo Gabriel Martínez So Itero, Jorge Daniel Ríos Arrañaga
- **Sistema de Identificación de Incendios en Espacios Interiores.** **C9**
Héctor Caballero Hernández, Marco Antonio Ramos Corchado, Vianney Muñoz Jiménez
- **Explorando el Potencial de la WBB Modificada: Un Estudio Preliminar en la evaluación del equilibrio y la prevención de caídas en entornos clínicos.** **C10**
Dra. Azucena E. Jimenez-Corona, Rigoberto Martinez Mendez, Estefania Olivares Palacios, Marisa Lara Bautista, Aislinn Joally Ramirez Mendez, Zeus Tlaltecutili Dominguez Vega
- **Fundamentos generales para la Protección de Software y Programas de Cómputo en México.** **C11**
Alfredo Aguilar Ruiz

Una introducción amable pero rigurosa al aprendizaje por refuerzo

A gently but rigorous introduction to reinforcement learning

Mauro Alejandro Montenegro Meza^{1*}
m.montenegro.meza@gmail.com

Rolando Menchaca Méndez¹
rolando.menchaca@gmail.com

Ricardo Menchaca Méndez¹
ricardo.menchaca@gmail.com

Centro de Investigación en Computación del IPN¹

RESUMEN

La interacción con el mundo es una de las principales formas en las que se genera el aprendizaje, pues es el medio por el cuál se obtiene información del entorno, y se experimentan relaciones causa-efecto. Esta idea de aprender mediante la interacción es un aspecto fundamental en muchas teorías del aprendizaje y, en este artículo abordaremos un enfoque computacional llamado *aprendizaje por refuerzo* (*Reinforcement Learning*, RL) además de construir de manera progresiva y sencilla sus bases matemáticas, así como los métodos principales de solución. Por último, mostramos aplicaciones y algoritmos que son relevantes en la industria e investigación.

Palabras clave: Aprendizaje por Refuerzo, Aprendizaje de Máquina, Inteligencia Artificial.

ABSTRACT

The interaction within the world constitutes one of the main ways in which learning is generated, as it is the way by which we obtain information from the environment and we experience cause-effect relationships. This idea of learning through interaction is a fundamental issue in many learning theories and, in this paper, we will address a computational approach called Reinforcement Learning (RL) and we will build in a progressive and simple way its mathematical basis, as well as its main solution methods. Lastly, applications and algorithms that are relevant in the industry and research are presented.

Keywords: Reinforcement Learning, Machine Learning, Artificial Intelligence.

1. INTRODUCCIÓN

En la literatura se denomina como “*aprendizaje por refuerzo*” (*Reinforcement Learning*, RL) tanto a una clase de problemas, a una clase de métodos que abordan dichos problemas, así como al campo que los estudia en el contexto del “*aprendizaje de máquina*” (*Machine Learning*, ML). El objetivo del RL no es generalizar situaciones no vistas durante el entrenamiento como en el *aprendizaje no supervisado*, ni encontrar estructuras escondidas entre los datos no etiquetados como en el *aprendizaje no supervisado*; en esencia, el objetivo principal del RL es aprender *políticas de control* que le permitan a un *agente* interactuar de manera exitosa en el *ambiente* que lo contiene. A pesar de las diferencias, las arquitecturas de RL pueden incorporar componentes de ambos paradigmas para mejorar su desempeño.

El estudio del RL tiene como fundamentos el comportamiento basado en estímulos y el desarrollo de la teoría del control. Sus bases principales son los estudios de comportamiento basado en estímulos de Pavlov, las *ecuaciones de Bellman*, y una versión estocástica del problema de control que modela la dinámica de problemas de toma de decisiones secuenciales por medio de *procesos de decisión de Markov* (*Markov Decision Process*, MDP).

2. PROCESO DE DECISIÓN DE MARKOV (MDP)

Como se muestra en la Figura 1, la arquitectura general de un sistema de RL está compuesta por un **agente** que interactúa con un **ambiente** a través de un conjunto de **acciones** que se seleccionan siguiendo una **política** de control π . El ambiente puede estar en cualquiera de los estados s definidos en su espacio de estados S y en cada instante t de tiempo discreto experimenta una transición a un nuevo estado $s' \in S$. La política de control $\pi : S \rightarrow A$ es una función que determina la acción que toma el agente dado un estado s . Cada que el agente realiza una acción a definida en el espacio de acciones A , se obtiene una señal de recompensa r_t .

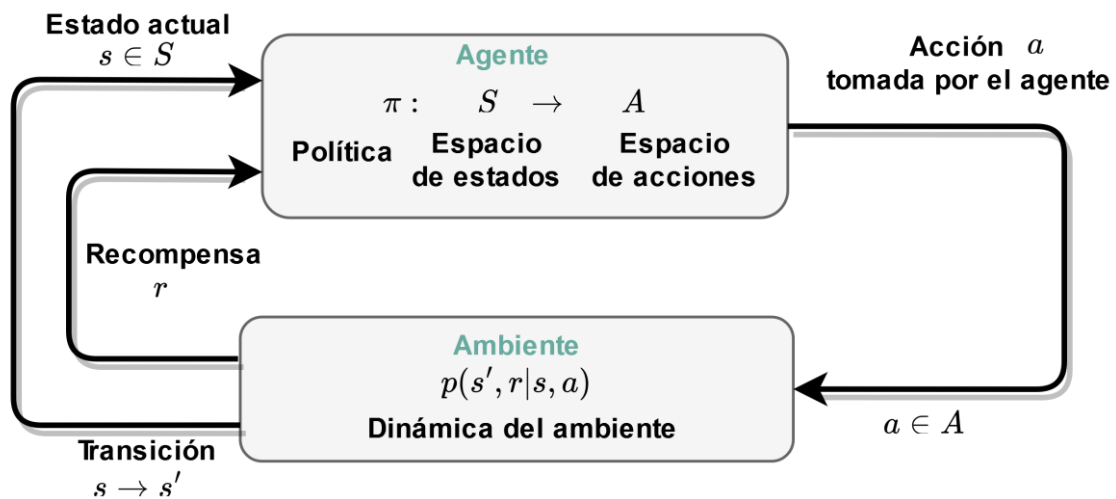


Figura 1. Arquitectura general de un sistema RL.

La dinámica de los sistemas de RL (ambiente + agente) se modela como un proceso de decisión de Markov (MDP) donde se asume que la probabilidad $p(s', r | s, a)$ de transitar del estado s al estado s' y obtener una recompensa r , solo depende del estado actual s y de la acción a tomada por el agente. Esto queda plasmado en la Ec.1 que nos dice que la probabilidad de que el siguiente estado en que estará el ambiente sea s_{t+1} y que se reciba una recompensa igual a r_{t+1} está condicionada al estado s_t en el que está el ambiente en el tiempo t , así como a la acción a_t tomada por el agente también el tiempo t . Esto significa que el sistema *no tiene memoria* y que por lo tanto la historia pasada no tiene injerencia en la transición actual. Es importante mencionar que, como se verá más adelante, los sistemas de RL son capaces de encontrar políticas de control óptimas aún y cuando se desconozca la dinámica del sistema.

$$p(s', r | s, a) = Pr\{r_{t+1} = r, s_{t+1} = s' | s_t, a_t\} \quad (1)$$

El objetivo de los algoritmos de entrenamiento de RL es encontrar una política de control π que maximice el valor esperado de la recompensa acumulada G_t que se define como la suma de las recompensas individuales r_t recibidas en cada instante de tiempo discreto t . Esto quiere decir que los sistemas de RL buscan seleccionar las acciones de control que maximicen la suma de las recompensas a largo plazo y no necesariamente la recompensa inmediata.

Para cuantificar la efectividad de una política π , se utiliza una *función de valor* de los estados $V^\pi(s)$, que precisamente determina el valor esperado de la recompensa que se obtendrá si el ambiente parte del estado s y el agente sigue la política de control π . Esta idea es expresada matemáticamente por medio de la Ec. 1 donde $E[G_t | s_t = s]$ se lee como el valor esperado de G_t dado que el ambiente está en el estado s_t . Las r_{t+k+1} son las recompensas que el agente obtendrá desde el tiempo $t + 1$ hasta el fin de la historia. La constante γ utilizada en la Ec. 2, llamada *factor de descuento*, sirve para atenuar la importancia de las recompensas futuras de las cuáles se tiene menos certeza. La Ec.2 se escribe en términos de un valor esperado debido a que tanto el ambiente como el agente pueden tener comportamientos aleatorios.

$$V^\pi(s) = E[G_t | s_t = s] = E\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\right] \quad (2)$$

De manera similar, la Ec. 3 define la función de valor de los pares estado-acción $Q^\pi(a, s)$ (*función estado-acción*) como el valor esperado de la recompensa acumulada al tomar la acción a en el

estado s al tiempo t , y posteriormente seguir la política π . En este caso, la función de estado-acción nos permite saber cuál será la recompensa acumulada esperada si es que tomamos una acción a_t en particular y no necesariamente la definida por π .

$$Q^\pi(s, a) = \mathbb{E}[G_t | s_t = s, a_t = a] = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a\right] \quad (3)$$

La Ec. 4 muestra la relación entre $Q^\pi(s, a)$ y $V^\pi(s)$ donde $\pi(a|s)$ es la probabilidad de tomar la acción a dado que el ambiente está en el estado s y se sigue la política π . La Ec. 4 nos indica que $V^\pi(s)$ puede verse como el valor esperado de las recompensas a futuro cuando se está en el estado s y se seleccionan acciones siguiendo la política π .

$$V^\pi(s) = \sum_a \pi(a|s) * Q^\pi(s, a) \quad (4)$$

Una forma de calcular de manera exacta los valores de $Q^\pi(s, a)$ y $V^\pi(s)$ es por medio de las *ecuaciones de Bellman*, Ec. 5 para la función de valor de los estados y Ec. 6 para la función de estado-acción. En ambas ecuaciones el valor esperado de la recompensa acumulada se calcula en términos del valor esperado de la recompensa r obtenida durante la transición de s a los posibles siguientes estados s' , más la formulación recursiva del valor esperado de la recompensa acumulada a partir de ese momento ($V^\pi(s')$ o $Q^\pi(s', a')$ según sea el caso), siguiendo la política π . En ambas ecuaciones, la sumatoria $\sum_{s', r}$ indica que se están considerando todos los posibles siguientes estados s' , así como todos los posibles valores de la recompensa r . Como se ilustra en la Figura 2, la Ec. 5 calcula el valor esperado de las recompensas que se obtendrán siguiendo la política π considerando todas las formas en que el ambiente transitará entre sus diferentes estados como resultado de las acciones de control tomadas por el agente. El caso de la Ec. 6 es análogo, sólo que se asume que se comienza seleccionando una acción a en particular.

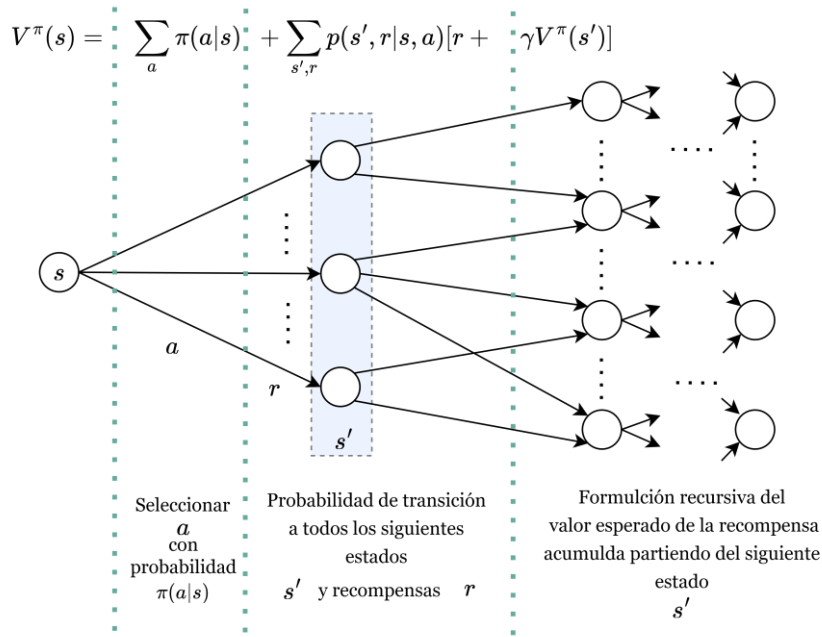


Figura 2. Las ecuaciones de Bellman consideran todas las formas en que puede evolucionar el ambiente dado que el agente sigue la política π .

$$V^\pi(s) = \sum_a \pi(a|s) * \sum_{s'} P_{s,s'}^a(r(s,a) + \gamma V^\pi(s')) \quad (5)$$

$$Q^\pi(s,a) = \sum_{s'} P_{s,s'}^a(r(s,a) + \gamma \sum_{a'} \pi(a'|s') Q^\pi(s',a')) \quad (6)$$

Una variante de las Ecs. de Bellman, conocidas como *Ecuaciones de optimalidad de Bellman*, pueden ser utilizadas para calcular la *política de control óptima* (π^*), es decir, aquella que maximiza la recompensa esperada de todos los estados en los que puede estar el ambiente. Más formalmente, decimos que una política π es mejor que otra π' si, y solamente si, $V^\pi(s) \geq V^{\pi'}(s)$ $\forall s \in S$. De esta forma, una política óptima se define como aquella que satisface $\pi^* \geq \pi$ para

todas las políticas π . Las funciones de valor óptimas $V^*(s)$ y $Q^*(s, a)$ se obtienen al seleccionar la acción a que maximice el valor esperado de la recompensa a largo plazo. Esto queda de manifiesto en las Ecs. 7 y 8 que en lugar de seguir una política arbitraria π , simplemente seleccionan la acción a que maximice el valor esperado.

$$V^*(s) = \max_a \sum_{s'} P_{s,s'}^a (r(s, a) + \gamma V^*(s')) \quad (7)$$

$$Q^*(s, a) = \sum_{s'} P_{s,s'}^a (r(s, a) + \gamma \max_{a'} Q^*(s', a')) \quad (8)$$

Buscar la solución a las ecuaciones de optimalidad de Bellman constituye uno de los problemas centrales del aprendizaje por refuerzo. Esto se debe a que a partir las funciones de valor óptimas es fácil derivar una política de control óptima. Simplemente se elige la acción que lleve al siguiente estado con mayor valor. En las secciones siguientes se presentan algunos de los métodos más representativos.

2. MÉTODOS TABULARES

En el método de programación dinámica (*Dynamic Programming*, DP) el valor de cada estado $V^\pi(s)$ se calcula a partir de la estimación actual del valor de sus estados vecinos. Dos estados s y s' son vecinos, si la probabilidad de que el ambiente transite en un solo paso de s a s' es mayor a cero, es decir si $p(s', r|s, a) > 0$ para cualquier acción a y recompensa r .

Todos los métodos tabulares que estudiaremos aquí operan siguiendo un proceso llamado “*Iteración general de políticas*” (*General Policy Iteration*, GPI) que consiste en iterativamente realizar los procesos de “*Evaluación de política*” (*Policy Evaluation*, PE) y “*Mejorado de política*” (*Policy Improvement*, PI). A grandes rasgos, el objetivo de PE es calcular la función de valor de la política actual y el de PI es obtener una mejor política a partir de dicha información.

El proceso de PE consiste en utilizar la Ec. 9 de manera iterativa para actualizar el valor actual de la función de valor. En la práctica, este proceso se realiza hasta que el cambio en los valores de los estados de dos iteraciones consecutivas sea menor a una cota ψ pre-establecida. Es importante notar que la Ec. 9 es una adecuación de la Ec. 5 utilizada como una regla de actualización. La demostración de convergencia de este procedimiento se basa en el hecho de que la Ec. 9 es un “*mapeo de contracción*” (*Contraction mapping*) y por lo tanto tiene un único punto fijo.

$$v_{k+1}(s) = \sum_a \pi(a|s) \sum_{s', r} p(s', r|s, a) [r + \gamma v^\pi(s')] \quad (9)$$

Por su parte, el paso de PI consiste en utilizar la ecuación de optimalidad de Bellman para definir una nueva política π' a partir de la función de valor de la política actual π . Como puede verse en la Ec. 10, π' es una versión voraz (*greedy*) de π que selecciona la acción a que maximice la recompensa esperada. Finalmente, observe que el caso en que π' es igual π , la Ec. 10 se convierte en la ecuación de optimalidad de Bellman y por lo tanto $\pi = \pi'$ es una política de control óptima. La Figura 3 ilustra la forma en que los procesos de PE y PI interactúan para moverse tanto en el espacio de las funciones de valor, como en el espacio de las políticas, para dirigirse a la política óptima π^* .

$$\pi'(s) = \arg \max_a \sum_{s', r} p(s', r|s, a) [r + \gamma v^\pi(s')] \quad (10)$$

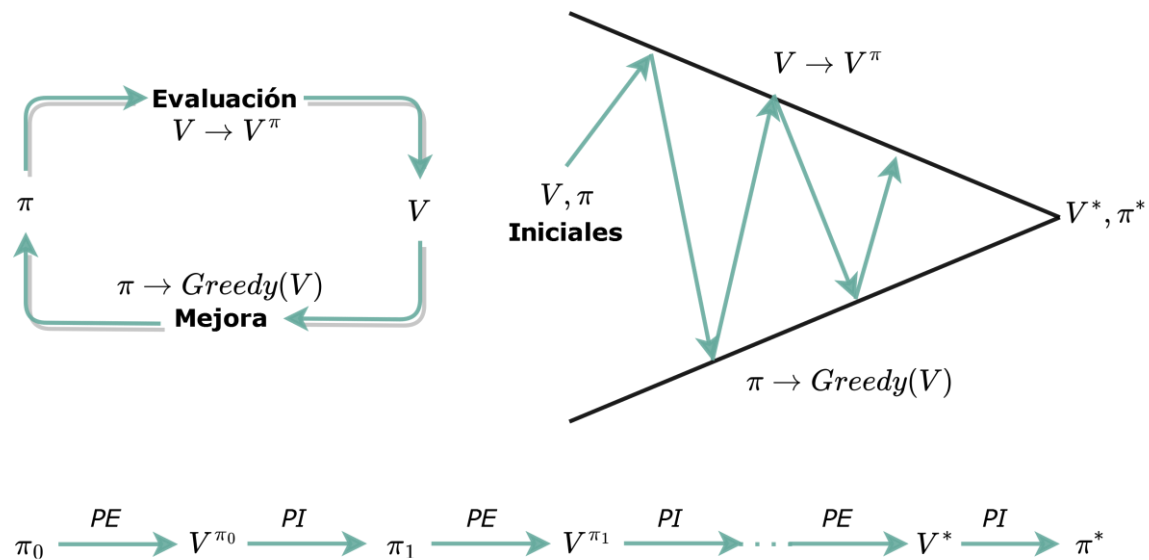


Figura 3. Convergencia a una política óptima en el proceso de Iteración General de Política.

Esta versión de PE tiene dos desventajas. Requiere conocer la función $p(s', r | s, a)$ que rige la dinámica del ambiente y puede requerir un número elevado de pasos para converger. Afortunadamente, las propiedades de convergencia de GPI a una política óptima no dependen de calcular de manera exacta el valor de las políticas. Partiendo de este hecho, el algoritmo de “Iteración de valor” (Value Iteration, VI) trunca la evaluación de la política a un solo paso y utiliza la Ec. 11 para combinar PE y PI en un único paso.

$$\begin{aligned}
 v_{k+1}(s) &= \max_a \mathbb{E}[R_{t+1} + \gamma v_k(S_{t+1} | S_t = s, A_t = a)] \\
 &= \max_a \sum_{s', r} p(s', r | s, a) [r + \gamma v_k(s')]
 \end{aligned} \tag{11}$$

A pesar de sus limitantes, los algoritmos clásicos de DP son importantes teóricamente y otros métodos tratan de emularlos pero de manera más eficiente.

Por su parte, los métodos *Monte Carlo* (MC) no requieren un modelo del ambiente y son capaces de aprender políticas por medio de la experiencia. De manera general, en los métodos MC el agente interactúa con el ambiente siguiendo una política π , almacenando los estados visitados, las acciones tomadas y las recompensas obtenidas en cada transición entre estados. En el caso de procesos episódicos, como una partida de Ajedrez, el proceso se repite varias veces. Al final,

se calculan estimados de la función $V^\pi(s)$ de cada estado visitado como un promedio ponderado basado en las recompensas acumuladas G_t experimentadas durante cada episodio.

En el caso de procesos episódicos (como una partida de Ajedrez), el proceso se repite N veces y, en cada iteración se calcula un promedio de las funciones. Existen dos algoritmos parecidos pero con propiedades teóricas particulares usados en los métodos MC: Cuando se promedia solamente la primera visita a cada estado (*algoritmo first visit*) o cuando se promedia cada vez que el estado es visitado en un episodio (*every visit*).

En los métodos MC la actualización de las funciones de valor se realiza mediante la Ec. 12 que calcula el valor de los estados como un promedio ponderado basado en la recompensa durante cada episodio (G_t).

$$V(S_t) = V(S_t) + \frac{1}{N} \alpha [G_t - V(S_t)] \quad (12)$$

Debido a que en MC no se conoce el modelo del ambiente, es necesario estimar explícitamente el valor de cada acción para poder elegir la mejor de ellas. De forma similar a DP, en MC el paso de mejora de política se realiza por medio de la Ec. 13, sólo que en este caso se trabaja con funciones $Q(s, a)$ que capturan el valor de cada par estado-acción.

$$\pi(s) = \arg \max_a q(s, a) \quad (13)$$

La principal ventaja de los métodos MC sobre los basados en DP es que son capaces de encontrar políticas de control optimizadas, aún y cuando no se tenga ningún conocimiento del ambiente. Lo anterior mediante un proceso en el cuál el agente experimenta las consecuencias, en términos de la recompensa obtenida, de tomar diferentes acciones en los estados que visita. Es importante mencionar que el cálculo de las políticas contempla la suma acumulada de las recompensas a largo plazo y no sólo las recompensas instantáneas de la acción inmediata. Desafortunadamente, para converger a una política óptima, los métodos MC requieren que el agente experimente un gran número de veces todos los pares estado-acción.

Una idea que combina los métodos MC y DP son los métodos de “*Diferencias temporales*” (*Temporal Difference*, TD). Los métodos TD aprenden directamente de la experiencia sin la necesidad de un modelo y actualizan sus estimados con base en otros estimados aprendidos anteriormente. El método TD más simple es conocido como TD(0) que, como se muestra en la Ec. 14, actualiza la función de valor $V(S_t)$ en cada transición del ambiente a partir de la recompensa inmediata R_{t+1} , más el estimado actual del valor del siguiente estado $V(S_{t+1})$.

$$V(S_t) = V(S_t) + \alpha[R_{t+1} + \gamma V(S_{t+1}) - V(S_t)] \quad (14)$$

Como MC, TD no requiere un modelo del ambiente, y como DP, TD aprovecha los estimados actuales de los valores de otros estados para actualizar inmediatamente el estimado del valor del estado actual.

Dentro de los métodos TD más populares están *SARSA* que utiliza la Ec. 15 para actualizar la función estado-acción, y *Q-Learning* que actualiza la función estado-acción acorde a la Ec. 16.

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha[R_{t+1} + \gamma Q(S_{t+1}, a) - Q(S_t, A_t)] \quad (15)$$

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)] \quad (16)$$

La diferencia entre *SARSA* y *Q-Learning* es sutil pero importante. *SARSA* es un método del tipo “*en la política*” (*on-policy*) que utiliza la misma política para evaluar estados-acciones y para tomar acciones de control, mientras que *Q-Learning* es un método del tipo “*fuera de la política*” (*off-policy*) que utiliza una política para evaluar estados-acciones y otra para tomar acciones de control. Normalmente, ambos algoritmos utilizan una política de control ϵ -voraz (ϵ -greedy) que selecciona una acción al azar con probabilidad ϵ , y que con una probabilidad de $1 - \epsilon$ se comporta de manera voraz seleccionando la mejor acción de acuerdo al estimado actual de la función de estado-acción. Para actualizar la función de estado-acción *SARSA* utiliza una política ϵ -greedy, mientras que *Q-Learning* utiliza una política voraz ($\epsilon = 0$). El objetivo de las políticas ϵ -voraz es permitir que el agente *explote* su conocimiento la mayor parte del tiempo (con probabilidad $1 - \epsilon$) pero que algunas veces (con probabilidad ϵ) *explore* acciones que tal vez no haya usado con anterioridad.

La Figura 4 presenta un resumen de los métodos tabulares mostrando sus pseudo códigos.

Inicializar $V(s), Q(s, a), \psi \geq 0, \pi$	
Programación Dinámica (VI)	Monte Carlo
Repetir $\delta \leftarrow 0$ Para cada estado $s \in S$ $v \leftarrow V(s)$ $V(s) \leftarrow \max_a \sum_{s', r} p(s', r s, a) [r + \gamma V(s')]$ $\delta \leftarrow \max(\delta, v - V(s))$ Hasta $\delta < \theta$	Repetir Infinitamente Generar episodio T siguiendo π Por cada estado $s \in T$ $G \leftarrow$ Retorno de primera incurrencia de s en T $R_{Total} \leftarrow G$ $V(s) \leftarrow \text{Promedio}(R_{Total})$ Por cada estado $s \in T$ $\pi(s) \leftarrow \max_a Q(s, a)$
TD (Q-Learning)	
Mejora de Política Evaluación de Política	Repetir para cada episodio T Por cada paso en T Escoger A a partir de S usando π (e. g. $\epsilon - greedy$) Tomar acción A , observar R, S' $Q(S, A) \leftarrow Q(S, A) + \alpha [R + \gamma \max_a Q(S', A) - Q(S, A)]$ $S \leftarrow S'$ Hasta que S sea estado terminal

Figura 4. Pseudocódigo de métodos tabulares.

3. MÉTODOS PARAMÉTRICOS

Los métodos paramétricos intentan solventar la falta de escalabilidad de los métodos tabulares por medio de *aproximaciones* y *generalización*. La idea consiste en utilizar modelos cuyos parámetros son optimizados para aproximar las funciones de valor (estado o estado-acción) a partir de las recompensas obtenidas al interactuar con el ambiente. Típicamente, el número de parámetros es mucho menor que el número de pares estado-acción. Una de las principales ventajas de estos modelos, como las redes neuronales artificiales (ANN), es su capacidad de generalización que permite tener estimaciones de valores para pares estado-acción aún y cuando el agente no los haya experimentado en el pasado.

En términos más concretos, las funciones aproximadas cuentan con un vector de parámetros $w \in \mathbb{R}^n$, donde $\bar{v}_\pi(s; \vec{w})$ es el valor aproximado del estado s dado un vector de parámetros $\vec{w}$, cuando se sigue la política π , es decir $\bar{v}_\pi(s; \vec{w}) \approx v_\pi(s)$.

El proceso de entrenamiento del modelo paramétrico consiste en optimizar el vector de parámetros $\vec{w}$ para minimizar el error observado en los ejemplos experimentados por el agente durante su interacción con el ambiente.

Por su simplicidad y efectividad práctica, uno de los métodos más usados para optimizar el vector de parámetros es el *Gradiente Descendente Estocástico* (*Stochastic Gradient Descent*, SGD), el cuál, ajusta los valores de $\vec{w}$ después de cada actualización de la recompensa acumulada G_t en la dirección que reduciría más el error $G_t - \bar{v}(S_t; \vec{w}_t)$ en ese ejemplo en particular.

La Ec. 17 muestra la regla de actualización de los pesos utilizando SGD, donde $\nabla \bar{v}(S_t; \vec{w}_t)$ denota el vector de derivadas parciales de $\bar{v}$ respecto a los parámetros $\vec{w}_t$, llamado gradiente de $\bar{v}$ respecto a $\vec{w}_t$.

$$\vec{w}_{t+1} = \vec{w}_t + \alpha [G_t - \bar{v}(S_t; \vec{w}_t)] \nabla \bar{v}(S_t; \vec{w}_t) \quad (17)$$

Uno de los métodos paramétricos más populares es una adaptación de Q-Learning, llamada Deep Q-Learning (DQN), que utiliza una red neuronal profunda (*Deep Neural Network*, DNN) como modelo de aproximación a las funciones estado-acción. La actualización de las funciones de estado-acción se realiza mediante la Ec. 16, pero sustituyendo los estimados de las funciones de estado-acción por un modelo con parámetros $\vec{w}$. La actualización de estos parámetros se realiza sustituyendo en la Ec. 17 el valor de G_t por la función objetivo $r + \gamma \max_a \bar{v}(S_{t+1}; \vec{w}_t)$.

En este caso, debido a que se utiliza el mismo vector de parámetros $\vec{w}$ para realizar la predicción $Q(s, a; \vec{w})$ y para evaluar la función *Q objetivo* $r + \gamma \max_a Q(s', a; \vec{w})$ existe el riesgo de crear oscilaciones durante el entrenamiento. Para minimizar el impacto de este problema se utiliza la Ec. 18 en donde se introduce una segunda DNN con parámetros $\vec{w}'$, llamada “red objetivo” (*target network*), para evaluar la función *Q objetivo*. Estos parámetros $\vec{w}'$ son actualizados periódicamente mediante copia de los parámetros $\vec{w}$ de la red principal.

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha [r + \gamma \max_{a'} Q(s', a; \vec{w}') - Q(s, a; \vec{w})] \quad (18)$$

Otro aspecto relevante a considerar en el contexto del uso de ANN en RL es el hecho de que la secuencia de recompensas experimentadas por un agente al interactuar con el ambiente están altamente correlacionadas, y por lo tanto, no deben ser usadas directamente para entrenar una ANN. Esto se debe a que los algoritmos de entrenamiento suponen que los ejemplos de entrenamiento son independientes e idénticamente distribuidos (*iid*). DQN aborda este problema por medio de una memoria donde almacena las experiencias pasadas, llamada *memoria de experiencia*, y que es muestreada aleatoriamente para obtener ejemplos de entrenamiento. Esta técnica es llamada *búfer de repetición* o *búfer de experiencia* y almacena gradualmente en un búfer las tóuplas de experiencias (S, A, R, S') al interactuar con el ambiente. Al acto de muestrear de

manera aleatoria pequeños conjuntos de tuplas del *búfer de experiencia* y utilizarlos en el aprendizaje se le llama *repetición de experiencia*. Además de romper éstas correlaciones entre las tuplas, el uso de ésta técnica permite aprender más acerca de ejemplos individuales al ser posible muestrearlos más de una vez y hacer mejor uso de ésta experiencia.

4. APLICACIONES EXITOSAS DE RL

En el ámbito científico, el primer gran éxito del RL fue *TD-Gammon* (Tesauro, 1995) que en esencia, es un programa computacional que utiliza una forma de TD ($TD - \gamma$) en el entrenamiento de una ANN para aprender a jugar de manera profesional Backgammon. En 2013, se presenta DQN para resolver problemas con un número elevado de estados como los juegos de Atari (Mnih et al., 2013). *AlphaGo* (Silver et al., 2016), desarrollado por DeepMind, aprende a jugar un juego altamente complejo como el Go mediante un *árbol de búsqueda monte carlo* y una red de política entrenada con *gradiente de política*. En 2020 se crea *AlphaZero* (Silver et al., 2018), una mejora de *AlphaGo*, que generaliza a juegos como el ajedrez y no utiliza información de expertos. Más recientemente, en 2022 *AlphaTensor* (Fawzi et al., 2022) fue capaz de descubrir algoritmos más eficientes a los anteriormente conocidos para multiplicación de matrices.

En la industria se encuentran aplicaciones que utilizan un esquema de GPI con restricciones y una arquitectura DQN en la optimización de sistemas de refrigeración para centros de datos (Luo et al., 2022). Por otra parte, Netflix utiliza SARSA y Q-Learning en sus nuevos sistemas de recomendación (Elahi, 2022), mientras que IBM los emplea en sus sistemas de predicción de mercado. En el área médica se utilizan políticas entrenadas mediante Q-Learning en la optimización de *regímenes de tratamiento dinámico* (DTR). En la Figura 5 se presenta una línea de tiempo con algunas de las aplicaciones exitosas de RL.

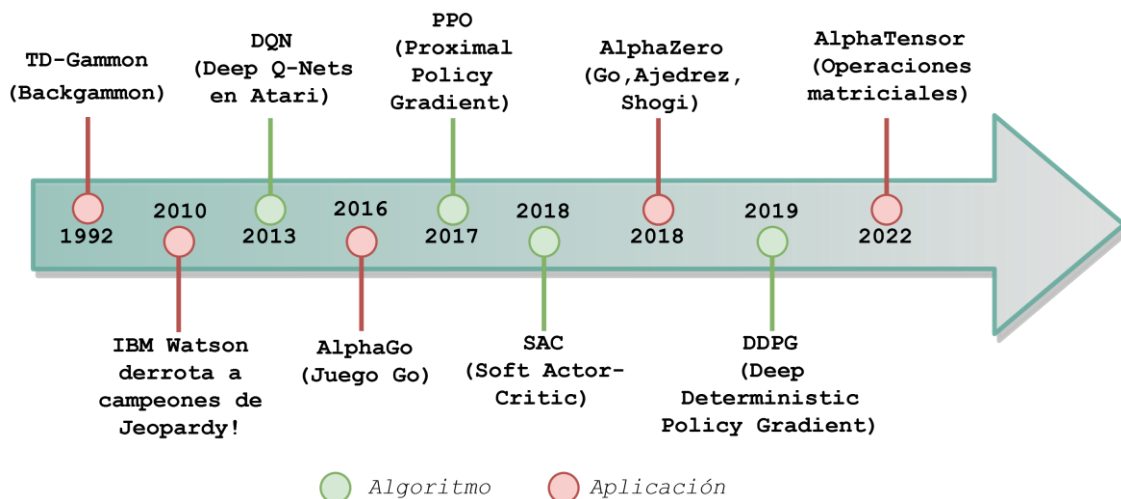


Figura 5. Línea de tiempo de trabajos sobresalientes en RL.

5. CONCLUSIONES

De forma similar a cómo los animales aprendemos, el proceso de aprendizaje en RL consiste en registrar las consecuencias que un agente experimenta al realizar acciones de control que afectan el ambiente que lo contiene. Conforme el agente acumula experiencias, los algoritmos de RL calculan políticas de control que maximizan el valor esperado de la cantidad de recompensas que el agente recibirá a lo largo del tiempo. En RL, los agentes se encuentran en una constante disyuntiva entre *explotar* el conocimiento adquirido que se encuentra codificado en la política de control actual para maximizar la recompensas y *explorar* acciones de control no utilizadas en el pasado con la esperanza de encontrar una mejor política de control.

Los algoritmos para calcular políticas de control se pueden clasificar en *métodos tabulares* (e.g., DP, MC, TD) en donde se almacenan las funciones de valor en tablas y en *métodos paramétricos* (e.g., DQN) que utilizan modelos parametrizados para aproximar dichas funciones. El progreso del RL en los últimos años ha sido notable y se espera que esta tendencia continúe tanto en el ámbito científico como el industrial. Al lector interesado en profundizar este tema se le recomiendan los libros de Sutton & Barto (Sutton & Barto, 2018) y de Bertsekas (Bertsekas, 2012).

6. REFERENCIAS BIBLIOGRÁFICAS

- Bertsekas, D. (2012). *Dynamic programming and optimal control: Volume i* (Vol. 1). Athena scientific.
- Elahi, E. (2022). *Reinforcement learning for budget constrained recommendations*. Retrieved 14 January 2023, from <https://netflixtechblog.com/reinforcement-learning-for-budget-constrained-recommendations-6cbc5263a32a>
- Fawzi, A., Balog, M., Huang, A., Hubert, T., Romera-Paredes, B., Barekatin, M., ... others (2022). Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, 610(7930), 47–53.
- Luo, J., Paduraru, C., Voicu, O., Chervonyi, Y., Munns, S., Li, J., ... others (2022). Controlling commercial cooling systems using reinforcement learning. *arXiv preprint arXiv:2211.07357*.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... others (2016). Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587), 484–489.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... others (2018). A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419), 1140–1144.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- Tesauro, G. (1995). Td-gammon: A self-teaching backgammon program. *Applications of neural networks*, 267–285.



Esta obra está bajo una licencia de Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Hacia una propuesta de rúbrica para la evaluación de la calidad de comentarios de código en Java

**Towards a rubric proposal for the evaluation of the quality of
code comments in Java**

Juan Carlos García Murillo¹
garciamjuancarlos14@gmail.com

RESUMEN

En un desarrollo de software la documentación juega un papel importante para conocer el origen, detalles, razón de ser, entre otros factores de un proyecto. Su utilidad es muy amplia y se extiende incluso como apoyo para que futuros desarrolladores puedan darle continuidad al sistema con menores dificultades. El código de un software es de los elementos más complejos de entender, ya que cada desarrollador posee una forma propia de programar. La manera por excelencia para hacer un código entendible es acompañarlo de buenos comentarios: completos y consistentes. En este artículo proponemos una rúbrica para evaluar la calidad de comentarios de código en Java. Dicha rúbrica está dividida por categorías según el tipo de comentario a evaluar: comentario de autoría, comentario de clase, comentario de método y comentario de fragmento de código, cada categoría establece un conjunto de reglas que deben cumplirse para declarar que un comentario es de calidad. El aumento en la legibilidad y mantenibilidad del código en proyectos de software que se encuentran actualmente en uso después de haber usado la rúbrica nos ayudará a validar su efectividad. Una vez comprobada esta eficacia ayudará a los equipos a poseer códigos correctamente comentados y a reducir costes en labores de entendimiento y mantenimiento del código pues este será más legible y descriptivo.

Palabras clave: Comentario de código, evaluación, Java, software, mantenimiento.

ABSTRACT

In a software development, the documentation plays an important role for knowing the origin, details and reason to be, among other factors of a project. Its utility is wide and is extended even as support for future developers can continue the development with less difficulty. The code of a software is one of the most complex elements to understand, since each developer has their own style of programming. The best way for making the code understandable is accompany it with good comments all of them complete and consistent. In this article, we propose a rubric to evaluate the quality of code comments in Java. This rubric is divided into categories according to the type of comment to be assessed: copyright comment, class comment, method comment, and code snippet comment. Each category defines a set of rules that must be met to declare a comment has quality. Through the increment of readability and maintainability of code in software projects that are currently in use after using the rubric will help us to validate its effectiveness. Once this effectiveness has been verified, it will help the teams to have correctly commented codes and to reduce costs in the work of understanding and maintaining the code, since it will be more readable and descriptive.

Keywords: Code comment, assessment, Java, software, maintenance.

1. INTRODUCCIÓN

Es común codificar software con miradas al corto plazo, o con un enfoque "egoísta", es decir, no se piensa muchas veces en que los integrantes de una empresa, institución o de un proyecto por más grande o pequeño que sea está sujeto a cambios en su equipo de trabajo.

Programar es un arte, no hay una forma cuadrada que obligue a todas las personas a realizar código siguiendo la misma lógica o forma. Esto complica la labor de entendimiento de un código ajeno, sobre todo al momento de darle mantenimiento a una pieza de instrucciones lógicas que no es de nuestra autoría. Incluso aun habiendo sido nosotros los desarrolladores de un código, después de un tiempo se vuelve complejo de entender.

En los lenguajes de programación existen los comentarios de código: un comentario de código es la interpretación y explicación del código. Estas descripciones utilizan el lenguaje natural para explicar la lógica y la función implementada detrás del código (Steidl et al., 2013) . En este contexto, los comentarios de código representan la principal fuente de documentación de un sistema de software y son, por lo tanto, fundamentales para el entendimiento del código tanto en la fase de desarrollo como en la fase de mantenimiento (Woodfield et al., 1981).

A pesar de que es posible comentar el código, muchos desarrolladores no realizan esta importante tarea, condenando a los futuros programadores a gastar mucho tiempo tratando de entender cómo funciona el software en cuestión; esto en parte es producto de que la ausencia de documentación en el código no impide la ejecución de los programas. Ahora bien, existen personas que sí comentan su código, pero a un nivel poco detallado o 'incompleto'.

En los proyectos de software el tiempo es una variable que debe manejarse con mucho cuidado ya que los retrasos pueden impactar negativamente al éxito, presupuesto y confianza del equipo de desarrollo. Los comentarios de mala calidad o la ausencia de ellos es un factor que puede orientar al proyecto a sufrir de este problema, por lo que es importante asegurar que en todo momento se cuente con documentación en el código y que la misma sea de calidad.

En este artículo definiremos una rúbrica para la evaluación de la calidad de comentarios de código en el lenguaje de programación Java, a fin de proveer a la comunidad una serie de reglas, que, de ser validadas, les permitirán reducir retrasos por entendimiento del código.

2. ESTADO DEL ARTE

Hirohisa Aman y sus colaboradores proponen evaluar cuantitativamente el valor de los comentarios de código aprovechando un modelo de procesamiento de lenguaje natural (NLP, por sus siglas en inglés) bien diseñado: Doc2Vec. En el artículo se considera un método o función como un documento y el contenido del mismo se expresa como un vector usando el modelo Doc2Vec. Después de esto, crean un nuevo vector con el mismo contenido del método, pero sin los comentarios de código. Luego, comparan los vectores obtenidos antes y después de la eliminación de los comentarios. A medida que los comentarios borrados brindan información más relevante, la diferencia entre los vectores se hace más grande. Sus resultados muestran que un método que tiene comentarios poco informativos es probable que sea propenso a sufrir modificaciones. (Aman et al., 2016).

Bai Yang y sus colaboradores recopilan investigaciones relevantes sobre los comentarios de código, incluyendo principalmente cuatro aspectos: generación automática de comentarios de código, consistencia de comentarios de código, clasificación de comentarios de código y evaluación de calidad de comentarios de código, que es el tema de nuestro interés. Para cada aspecto exponen técnicas que han sido utilizadas para abordar cada problema. En el ámbito de la evaluación de comentarios de código proporcionan una tabla comparativa que ilustra las investigaciones recopiladas sobre este tema exponiendo sus ventajas, desventajas, el método que utiliza cada una para realizar la evaluación, entre otras. Los resultados de esta investigación sirven como una guía para los inspectores de código para elegir las técnicas o herramientas adecuadas en materia de análisis y evaluación de comentarios de código (Bai et al., 2019).

En el estudio de Martina Iammarino y sus compañeros proponen evaluar la calidad de los comentarios de código en términos de coherencia con el código fuente, centrándose exclusivamente en su contenido. Presentan un enfoque basado en la técnica de *Topic Modeling*, y *Natural Language Processing*. El modelado de temas se refiere a modelos estadísticos utilizados para extraer temas que aparecen en una colección de documentos. Un tema es un conjunto de palabras similares, cada una asociada a un peso que expresa la probabilidad de que esté relacionada con el tema. El método utilizado para el modelado de temas en este trabajo tiene por nombre *Asignación Latente de Dirichlet*. Dicho método es una técnica no supervisada de minería de texto que es la encargada de asignar a cada palabra del documento un tema.

El proceso de evaluación propuesto consta de extraer los temas en el código fuente y en los comentarios de dicho código por separado y calcular la *Divergencia Kullback Leibler* que hay entre ellos a fin de conocer la coherencia que hay entre ambos. Los resultados del artículo muestran una similitud en la tendencia de distribución de temas, por lo tanto, casi todas las clases están asociadas a no más de tres temas (Iammarino et al., 2019).

Deze Wang y sus compañeros proponen tratar la evaluación de comentarios de código como un problema de clasificación donde el clasificador decide si la calidad del comentario es confiable o no. Se propone el uso de la red neuronal de entrada múltiple *DComment* para convertir en vectores al código y comentarios respectivamente basado en sus características. Dichos vectores son tratados por separado para finalmente introducirlos al clasificador que nos dirá si el comentario es confiable o no. Los resultados experimentales de la investigación muestran que el enfoque utilizado, en general, supera otras técnicas, tanto en su conjunto de datos etiquetado como en el conjunto de datos público, con una puntuación de 96.91% y 91.90% respectivamente. Usando los datos de entrenamiento y los de prueba de distintas fuentes, su enfoque aún puede lograr un rendimiento razonable, lo que demuestra su capacidad de generalización (Wang et al., 2019).

Pooja Rani en este estudio aborda el tema de la evaluación de comentarios desde tres perspectivas: qué preguntan los desarrolladores sobre prácticas de comentado de código, lo que estos mismos escriben en sus comentarios y cómo los investigadores los apoyan para evaluar la calidad de los comentarios. Para cada aspecto se realizó un análisis distinto a fin de generar hallazgos importantes.

En el primer aspecto el autor realizó un análisis de importantes foros de programadores tales como *StackOverflow* y *Quora*; como resultado de este proceso se presenta una taxonomía empíricamente validada de preguntas relacionadas con la convención de comentarios de varios foros comunitarios.

El segundo aspecto consta de un análisis de los comentarios de código que poseen proyectos de software realizados en *Java*, *Python* y *Pharo* donde se obtiene como resultado una taxonomía validada empíricamente que caracteriza los tipos de información que se encuentra en comentarios de clase.

En el tercer y último aspecto, el autor expone la planeación, objetivos, criterios de inclusión y exclusión, etc., de una futura revisión sistemática de literatura, donde se espera obtener como resultado hallazgos importantes para formular una taxonomía de calidad de comentarios que pueda ayudar a los investigadores y desarrolladores a identificar los atributos de calidad adecuados para cada tipo de comentario.

Los hallazgos preliminares de la investigación muestran que los desarrolladores incorporan varios tipos de información en los comentarios de clase en los lenguajes de programación. Aún así, enfrentan problemas para ubicar pautas relevantes en materia de escribir comentarios coherentes e informativos, verificar el cumplimiento de sus comentarios con las pautas y evaluar el estado general de la calidad de los comentarios (Rani, 2021).

3. RÚBRICA PARA LA EVALUACIÓN DE LA CALIDAD DE COMENTARIOS

Como pudimos ver, tres de los cinco artículos presentados en el estado del arte utilizan alguna técnica de análisis de datos o *machine learning* para realizar la evaluación de comentarios de código, dichos artículos son: *A Doc2Vec-Based Assessment of Comments and Its Application to Change-Prone Method Analysis* (Aman et al., 2016), *A Topic Modeling Approach To Evaluate The Comments Consistency To Source Code* (Iammarino et al., 2019) y *Deep Code-Comment Understanding and Assessment* (Wang et al., 2019). Esto convierte al uso de estas técnicas en el común denominador para abordar la evaluación de comentarios. Ante la necesidad de contar con equipos de cómputo con una alta capacidad de procesamiento para usar algoritmos de análisis de texto o de lenguaje natural, la rúbrica que será propuesta en este artículo, a diferencia de las investigaciones descritas con anterioridad, se compone de una serie de reglas, donde su cumplimiento será definido por un integrante del equipo de desarrollo (aunque esto a futuro se podría automatizar), eliminando así el requisito de contar con una alta y costosa capacidad de procesamiento cuando no se posea. Todo esto con el objetivo de ayudar a mejorar la calidad de los comentarios de código.

Java es uno de los lenguajes de programación más utilizados en la industria, ya sea para proyectos que llevan años existiendo o nuevos desarrollos. Dada la importante presencia de este lenguaje de programación en la tecnología actual, crece entonces la demanda de que el código hecho en Java posea comentarios de calidad que promuevan su legibilidad y mantenibilidad.

3.1 Categorías de comentarios en Java

En la Universidad de la prefectura de Okayama en Japón establecen cuatro categorías para dividir los comentarios que existen en Java, las cuales son: designación de derechos de autor, descripción simple de la clase, manual del programador para el método y explicación del fragmento de código (Aman et al., 2016), la figura 1 muestra un ejemplo de cada categoría. A continuación, se describe de una manera más amplia cada categoría:

- **Comentario de autoría:** Posee información relativa al o los creadores del código y sobre derechos de autor. Este tipo de comentarios en el código suelen estar al principio del archivo.
- **Comentario de clase:** Describe de manera general las funciones que desempeñan los métodos de la clase.
- **Comentario de método:** Describe la funcionalidad que desempeña el método, por ejemplo, el rol que funge cada parámetro, el procedimiento que se realiza dentro, los resultados que produce, así como el significado de cada uno de ellos. En palabras del autor “un manual para el programador sobre el método”.
- **Comentario de fragmento de código:** Describe el funcionamiento de una línea de código o un conjunto de líneas de código dentro de un método. Se suele hacer uso de esta categoría de comentarios cuando dentro de una función hay partes que son aún más complejas de entender a simple vista.



Figura 1. Categorías de comentarios

3.2 Reglas de la rúbrica

3.2.1 Reglas para la categoría de autoría

En esta categoría deberá situarse todo lo relativo a los derechos de autor. Según el diccionario de la real academia española podemos definir este concepto de la siguiente manera: “derecho que la ley reconoce al autor de una obra intelectual o artística para autorizar su reproducción y participar de los beneficios que esta genere”. (Real Academia Española, n.d.). Dicho así el código deberá estar protegido de acuerdos a las normas o licencia que prefieran el o los autores.

La declaración de derechos de autor debe situarse al inicio del archivo de código y contener los siguientes elementos: Palabra declarativa “*Copyright*”, año de creación, nombre del o los autores (persona física o empresa), oración “todos los derechos reservados”, licencia sobre la que se rige el código, enlace para consultar la licencia utilizada o descripción de la licencia dentro del mismo bloque de comentario.

3.2.2 Reglas para la categoría de clase

Esta categoría de comentarios debe componerse de los siguientes elementos: descripción general de las funcionalidades que desempeña la clase o lo que representa, nombre del o los autores y la versión. Para la versión y el autor es necesario utilizar las etiquetas *@author* y *@version* respectivamente (*How to Write Doc Comments for the Javadoc Tool*, n.d.). Es necesario mencionar que la descripción debe ser congruente con el código que contiene.

En Java existen variables a nivel de clase, en caso de existir, todas deberán estar comentadas con una descripción que permita conocer el rol que desempeña cada una de ellas.

Dentro de las clases podemos diferenciar tres tipos donde para cada uno la descripción de la clase debe seguir una cierta línea de especificación:

- **Clase matemática:** cuando la clase contiene operaciones matemáticas, transformaciones etc., su descripción requiere estar orientada a resumir estas funcionalidades, por ejemplo, una clase que realice operaciones estadísticas.
- **Clase de tabla:** cuando la clase representa un tipo de dato o tabla de una base de datos la descripción debe estar encaminada a comentar la información que almacena al momento de estar instanciada, por ejemplo, una clase que almacene la información de un “empleado”.
- **Clase combinada:** cuando la clase almacena datos y éstos mismos pueden ser manipulados dentro de la clase, la descripción debe definir qué datos almacena la clase y qué operaciones pueden aplicarse, por ejemplo, una clase “Arreglo”.

Las interfaces son un caso especial en Java, ya que no son una clase como tal, pero tienen un gran parecido, por lo tanto, las reglas a seguir para documentarlas son las mismas que fueron descritas con anterioridad.

3.2.3 Reglas para la categoría de método

La composición de un método en Java varía de acuerdo a lo que el programador necesita para codificar su funcionalidad, por lo tanto, la cantidad de reglas no es la misma para todos los casos. Toda función en Java comparte dos reglas: describir de manera detallada lo que realiza la función y que la descripción sea congruente con el código que contiene el método. Sin embargo, hay elementos adicionales que pueden estar presentes en la función y que su existencia requiere ampliar el comentario del método en cuestión. En la tabla 1 se describen las reglas a seguir cuando existe cada elemento.

Número de regla	Nombre del elemento	Descripción de la o las Regla(s)
1	Valor de retorno	Explicar lo que significa cada valor que regresa la función y bajo qué condiciones ocurre esto. El formato a usar es el siguiente: <i>@return [explicación]</i> .
2	Parámetro(s)	Explicar el rol que desempeña cada parámetro dentro de la función. El formato a usar es el siguiente: <i>@param [nombre del parámetro] [explicación]</i> .
3	Excepción(nes)	Explicar bajo qué condiciones ocurre cada excepción que lanza el método. El formato a usar es el siguiente: <i>@throws [nombre de la excepción] [explicación]</i> .
4	Anotación <i>@Deprecated</i>	Indicar desde qué versión el método ha sido declarado como obsoleto (<i>deprecado</i>) y el o los métodos que lo reemplazaron. El formato a usar es el siguiente: <i>@deprecated [versión en que se deprecó]. Reemplazado por [link #nombre_del_metodo]</i>
5	(<i>How and When to Deprecate APIs</i> , n.d.)	

Tabla 1. Reglas para los elementos variables en un método.

3.2.4 Reglas para la categoría de fragmento de código

La última categoría está compuesta por el código que se escribe dentro de los métodos. El objetivo aquí es conseguir un efecto en donde “la función nos platique a profundidad lo que realiza”, para conseguir esto es necesario establecer como regla que todo comentario, a excepción de las variables declaradas dentro de los métodos, deberán ser escritos en primera persona. Dicho esto, los fragmentos de código que se requieren comentar son los siguientes:

- Variables: describir el rol que desempeña.
- Condiciones: describir lo que evalúa.
- Ciclos: describir bajo qué condiciones el ciclo termina.
- Operaciones matemáticas con llamadas a funciones: describir lo que se obtiene a través de la llamada a otras funciones y el resultado que se consigue.

3.3 Uso de la rúbrica

Esta rúbrica será aplicable para asegurar la calidad de los comentarios de código cuando se utilice Java como lenguaje de programación. A través del cumplimiento de las reglas descritas se fomentará tener un código con comentarios altamente descriptivos lo que impactará positivamente en la legibilidad y por lo tanto en la mantenibilidad del código.

Deberán cumplirse todas las reglas definidas previamente, para esto se usará la tabla 2 que funge como el *checklist* de apoyo para la persona encargada de realizar la revisión. Los comentarios deberán analizarse por archivo de código, por lo tanto, no podrá dejarse incompleta la inspección de algún archivo y pasarse a otro. La rúbrica será utilizada por un integrante del equipo de desarrollo siempre y cuando no sea el autor del código a revisar y deberá llenar el *checklist* como se explica a continuación.

El *checklist* consta de un encabezado donde se deberá especificar el nombre del revisor, el nombre del archivo de código revisado y la fecha en que se realizó la revisión. posteriormente hay tres columnas:

- **Regla:** describe la regla que se tiene que cumplir,
- **¿Cumple?:** permite al revisor marcar con una “X” si la regla se cumple o dejar el espacio vacío en caso contrario. Existe la posibilidad de marcar la casilla con un N/A cuando la regla no aplique, por ejemplo, que la clase no contenga variables, que el método no posea parámetros, que no existan condiciones dentro del método, etc.
- **Observaciones:** tiene a fin permitir al inspector comentar lo que le falta al comentario para cumplir con la regla cuando sea el caso.

Se recomienda a los desarrolladores conocer la rúbrica previa a la escritura de sus comentarios para agilizar las revisiones de comentarios y reducir el incumplimiento de las reglas establecidas.

Una vez que se complete la revisión de los comentarios y la rúbrica esté debidamente llenada, se le hará llegar al equipo de desarrollo para que realice las correcciones pertinentes. Por ningún motivo deberán hacerse modificaciones al código si aún no se atienden las correcciones señaladas en la rúbrica. El revisor será el encargado de supervisar que los cambios solicitados sean aplicados.

Nombre del revisor:		Fecha:
Nombre del archivo de código:		
REGLA	¿CUMPLE?	Observaciones
COMENTARIOS DE AUTORÍA		
Se encuentra al inicio del archivo de código		
Contiene el formato: <i>Copyright [año de creación]. [Autor(es)] todos los derechos reservados</i>		
Nombra la licencia bajo la que se rige el código		
Contiene un enlace para consultar la licencia utilizada o describe la misma en el bloque de comentario		
COMENTARIOS DE CLASE O INTERFAZ		
Nombre de la clase o interfaz:		
Describe de manera general la clase		
La descripción es congruente con el código que contiene		
Cada variable de clase tiene una descripción del rol que desempeña		
Nombra al o los autores de la clase con el formato: <i>@author [nombre del autor]</i>		
Indica la versión de la clase con el formato: <i>@version [numero de versión]</i>		
<u>Indica la última fecha de actualización del comentario</u>		

COMENTARIO DE MÉTODO		
Firma del método:		
Describe detalladamente lo que realiza la función		
La descripción es congruente con el código que contiene el método		
Explica cada valor que regresa la función y bajo qué condiciones con el formato: <i>@return [explicación]</i>		
Explicar el rol que desempeña cada parámetro dentro de la función con el formato: <i>@param [nombre del parámetro] [explicación]</i>		
Indicar desde qué versión el método ha sido declarado como obsoleto (<i>deprecado</i>) y el o los métodos que lo reemplazaron con el formato: <i>@deprecated [versión en que se deprecó]. Reemplazado por [{@link #nombre_del_metodo}]</i>		
COMENTARIO DE FRAGMENTO DE CÓDIGO		
Las variables describen el rol que desempeñan en el método		
Las condiciones describen lo que evalúan		
La descripción de las condiciones está redactada en primera persona		
Los ciclos describen bajo qué condiciones terminan		
La descripción de los ciclos está redactada en primera persona		
Los comentarios de las operaciones aritméticas con llamadas a otras funciones describen lo que se obtiene a través de la llamada a los métodos externos y el resultado que se consigue de toda la operación		

Tabla 2. Checklist de la rúbrica para evaluar comentarios de código

4. DISCUSIÓN

Con el uso de esta rúbrica se espera que la industria de la ingeniería de software tenga una opción más a elegir para asegurar la calidad sus comentarios de código cuando se utilice Java como lenguaje de programación. Se pretende que los comentarios escritos bajo el conocimiento y uso de la rúbrica permitan a los programadores entender de una manera más rápida y sencilla el código, aumentar su legibilidad y así reducir gastos en mantenimiento. También se busca que los comentarios sean completamente congruentes con el código evitando así que al darle mantenimiento se olvide actualizar la documentación.

5. CONCLUSIONES

La investigación sobre la calidad de los comentarios en el código es un área que tiene mucho potencial de mejora. En la parte de la implementación dentro del ciclo de vida del desarrollo de software muchos de los esfuerzos se concentran en la elaboración de nuevos algoritmos, lenguajes de programación o tecnologías que permitan agilizar la tarea de codificar. Sin embargo, estos archivos con grandes cantidades de líneas de código además de cumplir exclusivamente con la funcionalidad para la que fueron creados deben aportar también valor para los desarrolladores cuya tarea sea darles mantenimiento. De aquí aparece la necesidad de contar con opciones que permitan realizar una evaluación a los comentarios escritos por los programadores a fin de asegurar su calidad. En esta investigación favorecemos a los proyectos que no cuentan con equipos de cómputo con altas capacidades de procesamiento ya que la rúbrica no requiere de estas facultades; de esta manera se habilita la opción de contar con un método de aseguramiento de calidad de comentarios de código sin costo computacional, que a futuro puede automatizarse generando así un costo que será mínimo comparado con otras técnicas dada su naturaleza.

Comentar el código es una tarea que erróneamente se deja a la posteridad o en el olvido generalmente por falta de tiempo. Miremos ahora esta labor como una inversión que será retornada al corto o largo plazo según la evolución del sistema en cuestión. Es necesario programar siempre bajo la premisa de que un buen código va acompañado de una buena documentación y si dicha documentación sigue un conjunto de reglas como las que fueron propuestas en este artículo, los futuros desarrolladores encontrarán comentarios con un alto grado de legibilidad, mantenibilidad, estandarizados y un mayor nivel de calidad.

6. REFERENCIAS

- Aman, H., Amasaki, S., Yokogawa, T., & Kawahara, M. (2016). *A Doc2Vec-Based Assessment of Comments and Its Application to Change-Prone Method Analysis*. <https://bit.ly/2Pb89Au>
- Bai, Y., Zhang, L., & Zhao, F. (2019). A survey on research of code comment. *ACM International Conference Proceeding Series*, 45–51. <https://doi.org/10.1145/3312662.3312710>
- How and When to Deprecate APIs*. (n.d.). Retrieved November 22, 2022, from <https://docs.oracle.com/javase/7/docs/technotes/guides/javadoc/deprecation/deprecation.html>
- How to Write Doc Comments for the Javadoc Tool*. (n.d.). Retrieved November 22, 2022, from <https://www.oracle.com/technical-resources/articles/java/javadoc-tool.html#tag>
- Iammarino, M., Aversano, L., Bernardi, M. L., & Cimitile, M. (2019). A Topic Modeling Approach To Evaluate The Comments Consistency To Source Code. In *IEEE Xplore*.
- Rani, P. (2021). Speculative Analysis for Quality Assessment of Code Comments. *Proceedings - International Conference on Software Engineering*, 299–303. <https://doi.org/10.1109/ICSE-Companion52605.2021.00132>
- Real Academia Española. (n.d.). *Derecho de autor*. Retrieved November 7, 2022, from <https://dle.rae.es/derecho#MYrPxAp>
- Steidl, D., Hummel, B., & Juergens, E. (2013). Quality Analysis of Source Code Comments. In *2013 IEEE 21st International Conference on Program Comprehension* (pp. 83–92). IEEE.
- Wang, D., Guo, Y., Dong, W., Wang, Z., Liu, H., & Li, S. (2019). Deep Code-Comment Understanding and Assessment. *IEEE Access*, 7, 174200–174209. <https://doi.org/10.1109/ACCESS.2019.2957424>
- Woodfield, S. N., Dunsnore, H. E., & Shen, V. Y. (1981). *The Effect of Modularization and Comments on Program Comprehension*.



Esta obra está bajo una licencia de Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Recibido 09 Nov. 2022

ReCIBE, Año 12 No. 1, May. 2023

Aceptado 14 Nov. 2022

Usability testing for therapist module on virtual environment for poststroke motor rehabilitation of upper limb

Prueba de usabilidad al módulo del terapeuta en un ambiente virtual para la rehabilitación motriz pos-ictus del miembro superior

Guadalupe Toledo Toledo¹
gtoledo_1207@hotmail.com

Aguilar Acevedo Francisco¹
aguilar.afco@sandunga.unistmo.edu.mx

Nieva García Omar Santiago¹
omarng@sandunga.unistmo.edu.mx

Daniel Pacheco Bautista¹
dpachecob@bianni.unistmo.edu.mx

Bezares Molina Francisco Gabriel¹
gabriel_m0104@live.com

Marleydy Acevedo Gómez¹
marscaru@gmail.com

Departamento de Ingeniería en Computación; Universidad del Istmo¹

Resumen

El presente artículo describe la implementación del protocolo Pensando en Voz Alta en su modalidad de Intervención Activa como fase de evaluación de un ambiente de realidad virtual, que apoya al terapeuta durante las sesiones de rehabilitación con pacientes que perdieron movilidad en el miembro superior debido a un accidente cerebrovascular, dicha estrategia es de bajo costo, y los resultados cualitativos permiten comprender en una primera instancia la calidad del software. La prueba estuvo dirigida hacia terapeutas especialistas en terapia física, quienes fungen como usuarios principales del entorno virtual y quienes configurarán las terapias que sus futuros pacientes realizarán. Como resultado de la prueba, se lograron recolectar 19 problemas de usabilidad clasificadas por gravedad y prioridad de corrección, considerados dos de Alta Prioridad, cinco de Mediana Prioridad y doce de Baja Prioridad, estos últimos orientados a pequeñas modificaciones a nivel de interfaz. De esta forma, se consiguió un primer acercamiento a la expectativa del usuario final respecto al ambiente virtual evaluado, con el fin de obtener una nueva iteración del desarrollo del software que permita reducir o anular dichos problemas, lo que sugiere por un lado los mecanismos para el refinamiento inmediato del software, y por otro, un antecedente que permita adecuar el procedimiento de evaluación a otros software destinados a la rehabilitación.

Palabras-clave: Accidente cerebrovascular, Pensando en voz alta, Pruebas de usabilidad, Realidad virtual.

Abstract

This article describes the implementation of Thinking Aloud protocol in its Active Intervention modality as an evaluation phase of a virtual reality environment. It supports the therapist during rehabilitation sessions with patients who have lost mobility in the upper limb due to a stroke. This strategy is low cost, and the qualitative results allow us to understand the quality of the software in the first instance. The test was aimed at therapists specializing in physical therapy, who serve as the main users of the virtual environment and who will configure the therapies that their future patients will carry out. As a result of the test, it was possible to collect 19 usability problems classified by gravity and correction priority, considered two of High Priority, five of Medium Priority and twelve of Low Priority, the latter oriented to small modifications at the interface level. In this way, a first approach to the end user's expectations regarding the evaluated virtual environment was achieved, in order to obtain a new iteration of software development that allows reducing or canceling problems, which suggests, on the one hand, the mechanisms for immediate refinement of the software, and on the other, a background that allows the evaluation procedure to be adapted to other software intended for rehabilitation.

Keyword: Stroke, Thinking aloud, Usability testing, Virtual reality.

1. Introducción

La rehabilitación pos-ictus busca ayudar al paciente a recuperar habilidades que perdió o vio disminuidas, a ser más independiente, y en general, a tener la mejor calidad de vida posible. El proceso de rehabilitación se aborda desde diferentes enfoques terapéuticos, ya sea con ejercicios intensivos de motricidad en el caso de la fisioterapia, o con actividades de la vida diaria (AVD) como en la terapia ocupacional (TO). Al respecto, una alternativa tecnológica a la que se ha recurrido es la Realidad Virtual (RV), la cual presume ser una oportunidad para integrar terapias que motiven al paciente a realizar las tareas virtuales (Bermudez i Badia, Fluët, Llorens, y Deutsch, 2016). No obstante, prevalece la interrogante de cómo garantizar que estas intervenciones de rehabilitación virtual satisfagan las necesidades específicas de los usuarios, siendo la incorporación de la perspectiva del usuario en el diseño, la identificación y atención a los problemas de usabilidad, dos de las líneas de estudio, que buscan hacer estos entornos virtuales “más útiles” como herramientas de apoyo en programas de rehabilitación (Aguilar, Pacheco, Acevedo, Toledo, y Nieva, 2022).

En lo que concierne a la perspectiva del usuario, un marco de trabajo usado en el desarrollo de estos ambientes virtuales ha sido el Diseño Centrado en el Usuario (DCU), el cual permite su aplicación en un gran abanico de contextos, siempre y cuando se tenga en claro quién toma el rol del usuario final, que en el contexto particular de la rehabilitación virtual puede ser el paciente y/o el terapeuta. Sobre este último enfoque, estudios como los de Korn y Tietz (2017), señalan la necesidad de que el terapeuta elija la combinación óptima de movimientos, de modo que la inmersión en la tarea virtual, no sea una distracción, mientras en Hung, Huang, Chen, y Chu (2016), señalan la preocupación que existe por parte de los pacientes sobre la correcta realización de sus ejercicios, lo cual otorga gran relevancia a las sugerencias y correcciones que el terapeuta realiza durante la ejecución de los mismos.

En lo respecta a la aceptación del producto por parte del usuario final, se ha recurrido a tanto a encuestas y cuestionarios (Sáenz, García, y Méndez, 2015), como al apego a reglas y estilos de diseño dado por expertos en el área (Shiratuuddin et al., 2017). En este punto, investigaciones como la presentada en este artículo se centran en explorar alternativas de evaluación y por consiguiente variaciones en el proceso metodológico de su aplicación, que permitan identificar problemas de usabilidad, tal es el caso de la aplicación del protocolo Pensando en Voz Alta (*Thinking Aloud*).

En particular, el protocolo Pensando en Voz Alta ha sido un instrumento usado ampliamente en el estudio del proceso cognitivo de las personas (Jääskeläinen, 2010), que evalúa mediante la recolección de datos cualitativos (estados de ánimo mientras se realiza la prueba, gustos/disgustos, entre otros) o cuantitativos (cantidad de problemas identificados, tiempo ejercido para la realización de las tareas, número de tareas completadas, entre otros), el nivel de éxito que se obtiene en la interacción con el objeto de estudio, que para los fines de esta investigación lo representa el software presentado en (Aguilar et al., 2022), el cual es un ambiente virtual para la rehabilitación motriz pos-ictus, diseñado bajo la filosofía del DCU, y destinado a mejorar la movilidad de la articulación del codo a través de actividades virtuales relacionadas con comer, el aseo personal y vestirse.

2. Protocolo Pensando en Voz Alta

La prueba Pensando en Voz Alta implica la intervención de dos personajes cruciales: el usuario, quien realizará una serie de tareas previamente definidas; y el experimentador o moderador quien realizará intervenciones durante el desarrollo de la estrategia para fomentar la recolección de datos (Lauesen, 2005).

El protocolo Pensando en Voz Alta es considerado un método de introspección, debido a que como se menciona en Van Someren, Barnard, y Sandberg (1994) citado en Ávila, Bianchetti, y González (2017), se puede "...observar los eventos que suceden en el subconsciente de una persona a partir de lo que puede observarse en los eventos que realiza en el mundo externo". Esto quiere decir que no se espera que el usuario narre lo que hace sino propiamente por qué lo hace y evitar en lo posible grandes periodos de silencio, lo que puede provocar pérdida de datos valiosos dentro del proceso cognitivo. A su vez, este protocolo ofrece la ventaja de entender la toma de decisión, de resolución de problemas y de pensamiento crítico del usuario (Ávila et al., 2017). Según Krahmer y Ummelen (2004), el protocolo Pensando en Voz Alta se concentra en tres objetivos principales: i) para encontrar evidencia de modelos y/o teorías del proceso cognitivo, ii) para descubrir y comprender patrones generales de comportamiento por parte de los usuarios en la interacción con documentos o aplicaciones con el fin de crearse bases científicas que representen reglas de diseño, y finalmente iii) para evaluar documentos o aplicaciones con el propósito de especificar problemas trascendentales y revisarlos.

En este sentido, en Boren y Ramey (2000) se presenta un estudio que destaca el tercer objetivo, concerniente a las pruebas de usabilidad, siendo el principal foco de atención el analizar las interacciones que el usuario realiza con el software, es decir, el usuario no se lleva toda la primicia, ya que el objeto de interés es el software y lo bien que está implementado para cumplir con sus funciones, por lo que el moderador encuentra una participación activa, dirigida y estratégicamente intervenida para identificar todos los aspectos que puedan impedir el éxito de la implantación de un determinado software.

En este sentido, en la variante del protocolo Pensando en Voz Alta conocida como Intervención Activa (Dumas y Redish, 1999; Olmsted-Hawala, Murphy, Hawala, y Ashenfelter, 2010), el moderador prueba obtener o concebir el modelo mental o propiamente el proceso de pensamiento del usuario para saber cómo ellos realizan una determinada actividad a través de la intervención activa (narrativa, conversación, impresiones) dentro de la prueba. Demostrando que la técnica de *coaching* mejora el desempeño de los usuarios e incrementa su satisfacción respecto a las otras técnicas.

Lo anterior permite confirmar la reflexión de Nørgaard y Hornbæk (2006) al mencionar que el moderador previo a la aplicación de la prueba ya puede intuir los tipos de problemas que pueden estar presentes durante la realización de la prueba y con ello, conducir al usuario hacia la confirmación de sus sospechas, por lo que aprueba la intervención oportuna y conducida del moderador.

Los argumentos expuestos sobre el protocolo Pensando en Voz Alta han dado paso a un interés por describir el procedimiento experimental que implica el uso del protocolo en su modalidad de Intervención Activa, siguiendo la propuesta de aplicación y reporte de resultados dada por Lauesen (2005), con la finalidad de obtener al cierre del protocolo una lista de problemas de usabilidad del objeto de estudio (ambiente virtual), clasificados por gravedad y prioridad de solución, ofreciendo como aportación una descripción detallada de los resultados obtenidos, así como las sugerencias inmediatas que darán solución a los problemas de usabilidad con la intención de proceder a realizar las mejoras del software, dentro del marco del DCU bajo el cual se desarrolló el ambiente virtual bajo estudio.

Es importante señalar, que al momento no se ha identificado trabajos relacionados en el campo de la rehabilitación virtual que haga uso del protocolo Pensando en Voz Alta como alternativa para medir e identificar aspectos a mejorar en la usabilidad de estos entornos virtuales.

3. Materiales y equipo

El objeto de estudio, lo constituye un sistema de RV de tipo no inmersivo, que emplea como periférico de interacción al sensor Kinect y como software de desarrollo al motor gráfico Unity. El diseño del software sigue la filosofía del DCU bajo el rol del terapeuta como usuario final, teniendo como propósitos ser una herramienta de apoyo que permita administrar (Figura 1), configurar (Figura 2) y realizar (Figura 3) las sesiones de terapia, teniendo como fundamento la retroalimentación visual como medio para motivar al paciente a realizar tareas virtuales relacionadas a cubrir la necesidad de ocuparse de sí mismo. Al tratarse de un software asistido (por el terapeuta), el ambiente virtual se conforma de dos módulos, el primero asociado a las tareas de administración y configuración denominado módulo del terapeuta, y el segundo relacionado con la realización de las tareas virtuales llamado módulo del paciente.



Figura 1. Registro de paciente. Fuente: (Acevedo, 2018)



Figura 2. Configuración de la terapia. Fuente: (Acevedo, 2018)



Figura 3. Tarea de cortar carne en escenario Comedor. Fuente: (Acevedo, 2018)

Respecto a la evaluación del software bajo mención, ésta la han concretado al evaluar el módulo del paciente (Aguilar et al., 2022), con participantes con y sin diagnóstico de ictus, mediante pruebas de usabilidad tipo *testing* a través de una prueba SUS (Bangor, Kortum, y Miller, 2008), un cuestionario de estado de flujo (Yoshida et al., 2013) y entrevistas para valorar si su interpretación del contenido e interacción con el ambiente virtual fueron de utilidad y correcta comprensión, así como para conocer sus opiniones en la efectividad de la terapia con apoyo de la tecnología de RV. Sin embargo, este primer acercamiento no ha permitido indagar la usabilidad del software desde la perspectiva del terapeuta, quien funge como usuario final y para quien fue diseñado el ambiente virtual. Así, en este trabajo se considera la aplicación de la técnica Pensando en Voz Alta bajo la modalidad de Intervención Activa, para la descripción de los problemas de usabilidad detectados, con el objetivo de contar con un mecanismo que permita valorar las opiniones de usuarios potenciales, terapeutas, quienes no participaron en la formulación de los requerimientos del software, pero se ajustan al perfil del usuario final que pudiese emplear este entorno.

De forma particular, según (Acevedo, 2018) el entorno de RV bajo estudio cumple con los siguientes requisitos funcionales:

- Acceso como administrador del sistema por el especialista a través de una clave.
- Registro de pacientes recabando datos generales y valoración con base en las escalas de Barthel, Ashworth y Daniels (véase Figura 1).
- Búsqueda de paciente.
- Revisión del historial de terapias de cada paciente.
- Eliminar el registro e histórico de pacientes.
- Acceso a sesiones específicas por paciente a través de su clave de identificación.
- Diseño de escenas de AVD implementadas mediante el uso del sensor Kinect para la identificación del movimiento de flexión del codo.
- El sistema tiene la capacidad de medir el ángulo de flexión del paciente al momento en que éste se encuentre realizando el ejercicio asignado, de tal manera que se puede verificar con exactitud si se está logrando el ángulo de flexión indicado por el especialista en la configuración previa al inicio de la sesión (véase Figura 2), independientemente de que el codo lesionado sea el derecho o izquierdo.
- El sistema retroalimenta a cada instante a través de paneles informativos sobre el ángulo de flexión logrado por el paciente al realizar la tarea virtual, un cronómetro en cuenta regresiva y el número de repeticiones realizadas (véase Figura 3).
- Genera un informe en formato Excel sobre los resultados de la sesión realizada con el resumen de estas variables

Para efectuar el protocolo Pensando en Voz Alta, se requiere como entrada principal usuarios finales o potenciales que se ajusten al perfil del terapeuta que participó en la formulación de software, así como las tareas de usabilidad que definirán las actividades que los participantes intentarán realizar durante su exploración en el ambiente virtual, obteniéndose como resultado una lista de problemas de usabilidad clasificados por su gravedad. Dicho protocolo se describe claramente en (Lauesen, 2005) y se expresa en formato de pseudocódigo en (Toledo, Rosales, Hernandez, y Arellano, 2015). Para ello, se estableció contacto con terapeutas que se ajustaran al perfil de usuario con título profesional y/o especialidad en Medicina Física y/o Rehabilitación, con un mínimo de 2 años de experiencia, y sin experiencia en el uso de un software similar. Previo al comienzo de la prueba, se procedió a entrevistar individualmente a cada terapeuta (Usuario potencial) a través de las interrogantes plasmadas en la Tabla 1, con la intención de validar su familiaridad con los conceptos propios del contexto de uso del software y por tanto con la configuración de la terapia.

	Usuario 1	Usuario 2	Usuario 3
Edad	26	26	25
Título profesional	Lic. en Fisioterapia	Lic. en Fisioterapia	Lic. en Fisioterapia
Especialidad	No	Rehabilitación	No
Experiencia	2 años	4 años	3 años
¿Ha usado un software parecido?	No	No	No
¿Conoce las escalas de Barthel, Ashworth, y Daniels?	Si	Si	Si

Tabla 1. Resultado de encuesta inicial a participantes Fuente: Autoría propia.

Posteriormente, se explicó la mecánica de la prueba a cada uno de los participantes, en la cual tenían frente de si (por primera vez) a la interfaz de usuario del ambiente virtual a evaluar, acompañados del facilitador de la prueba quien fungió como el paciente con la intención de hacer una demostración a cada participante sobre el tipo de retroalimentación visual que la interfaz ofrecía con ayuda del sensor Kinect, para esto, se le pidió autorización a cada especialista de grabar la sesión de prueba, y así recabar con mayor detalle algunos problemas de usabilidad que no fuesen identificados de inmediato. En la Tabla 2 se describen las tareas de usabilidad aplicadas, que cubren en su totalidad las funcionalidades ofrecidas para el módulo del terapeuta.

No .	Descripción
1	Por la tarde, se presenta un Nuevo caso a tu consultorio, es un niño de 13 años que lleva por nombre Fernando Pérez, el cual sufrió un accidente cerebro-vascular hace cinco semanas, que afectó en gran medida su codo izquierdo, identificas una dependencia severa al realizar actividades de la vida diaria, tales como lavarse los dientes o tomar un vaso; a su vez presenta una flexión máxima del codo de 100°, así como contracción muscular débil, diagnosticando a su vez, hipertensión moderada. Llevamos de un mes realizando terapia con otros especialistas. ¿Qué debes hacer para darlo de alta en el sistema?
2	Deseas revisar que los datos de Fernando Pérez quedaron correctamente almacenados, ¿Qué harías?
3	Margarita Sibaja, una de tus primeras pacientes con lesión en el codo derecho ha finalizado por fin sus terapias, las cuales realizó con ayuda del software, por lo que su registro dentro de éste ya no es necesario, ¿Qué tendrías que hacer?
4	Fernando Pérez comenzó su terapia apoyado del software, para ello decides configurar su prueba seleccionando como actividad de la vida diaria la acción de comer un sándwich, dicha configuración considerará una sesión de 3 minutos en los que ejercitará un ángulo de flexión de 120°, ¿Qué tienes que hacer?
5	Mientras Fernando Pérez realiza la terapia, te quedas cerca de él para estudiar su desempeño sin intervenir, podrías indicarme durante el minuto 2 ¿Qué ángulo logró alcanzar el paciente?, ¿Cuántas repeticiones logró realizar en la sesión?, cierras tu evaluación anexando al expediente tu punto de vista respecto a la terapia y creando dicho expediente en digital.
6	Has terminado de utilizar el software por el día de hoy y necesitas finalizar tu sesión activa, como lo harías dentro del sistema.

7	A los 3 días Fernando Pérez vuelve para su siguiente terapia, optas en esta ocasión por la dinámica de cepillado de dientes, configurando la sesión a un minuto con ejercicios de flexión en un ángulo de 125° ¿Cómo se ejecuta esto dentro del software?
8	Durante la ejecución de esta segunda sesión para Fernando Pérez te das el tiempo de analizar el comportamiento de su flexión, ¿Qué información te ofrece la pantalla respecto a Fernando Pérez?
9	Franco Yavhe es de tus pacientes más antiguos y ha venido a visitarte por una consulta, para determinar su próxima terapia necesitas revisar su progreso en las últimas dos sesiones, ¿Qué tendrías que hacer?
10	Al mes de proceso de recuperación de Fernando Pérez te visita, para consultar su progreso, sin embargo, de todos los pacientes que has atendido en ese tiempo, no puedes recordar sus últimos resultados, ¿Qué harías?

Tabla 2. Tareas de usabilidad. Fuente: Autoría propia.

4. Métodos experimentales

Se procedió a aplicar la prueba Pensando en Voz Alta, con los tres especialistas en el área y obtener retroalimentación inmediata sobre las expectativas en cuanto al diseño de la interfaz de usuario y el nivel de acoplamiento de su modelo mental respecto al programador del software. De acuerdo con (Nielsen, 1993), este número de usuarios es suficiente para obtener tantos problemas de usabilidad como estos existan, debido a que al aumentar el número de usuarios solo se obtendría la identificación de los mismos problemas y el coste de ejecución se elevaría innecesariamente.

Tal como indica Nielsen (1993; 2012) esta prueba es un proceso muy „barato” para obtener información pese a su extenuante ejercicio de recopilación y desglose de los resultados, no obstante, dichos resultados pueden ser útiles para analizar estadísticamente el grado de éxito de la interfaz en términos del número de problemas de usabilidad identificados, su gravedad, y además de ello, identificar nuevos requerimientos que pudiesen ayudar a una interacción exitosa con la herramienta.

Para ello, se procedió a realizar una estimación del plan de prueba (véase Tabla 3 y 4), obteniéndose un costo de ejecución de 10hrs. 54 min para la aplicación de la prueba de usabilidad a los tres terapeutas, que, si bien resulta un costo alto de tiempo, el impacto económico requerido es prácticamente nulo y los resultados pueden ser muy útiles para establecer un prototipo cada vez más agradable y satisfactorio.

Acción	Tiempo
Bienvenida e Introducción	5 min
Tareas	30 min
Reunión de conclusión	3 min
Escribir reporte	120 min
Tiempo total (por usuario)	158 min

Tabla 3. Costo de ejecución por usuario. Fuente: Autoría propia

Acción	Tiempo
Tiempo por tres usuarios	474 min
Escribir reporte de prueba	180 min
Tiempo total (por usuario)	10 hrs 54 min

Tabla 4. Costo de ejecución del plan de pruebas. Fuente: Autoría propia.

5. Discusión de resultados

Como resultado de la aplicación y recolección de respuestas por parte de los usuarios, se identificaron 19 problemas de usabilidad, mostrados en la Tabla 5.

No.	Descripción
1	El botón de salir no pide una confirmación (“¿Desea salir?”)
2	No aparece en la pantalla si se debe realizar un movimiento de flexión de codo, rotación de hombro o ambos.
3	No se puede “dar de alta” a un paciente.
4	La opción “eliminar” no debe aparecer en el sistema.
5	La búsqueda de los pacientes es tardada.
6	La función “volver” en la búsqueda de pacientes regresa a la pantalla de “login”.
7	El usuario no pudo “iniciar sesión” o “ver los datos” de algún paciente.
8	La última cita debe aparecer hasta arriba del historial.
9	La frase “Tiempo en terapias” da a entender “cuánto tiempo estará en terapias el paciente y no cuánto tiempo lleva ya en terapia.
10	No aparecen instrucciones de cómo realizar las AVD.
11	Al registrar el usuario no aparece un apartado para registrar qué “enfermedad” originó suproblema.
12	El calendario para establecer una fecha no se le entiende.
13	No aparece un menú cuando el “admin” (administrador) inicia sesión.
14	El encabezado del apartado del paciente no se le entiende (no es legible, formateo difícil de leer).
15	La tecla “enter” no funciona cuando pone el id de usuario.
16	No se puede ver quién fue el último terapeuta que atendió al paciente.
17	La imagen donde se ve al paciente realizar la tarea, es muy pequeña.
P18	En el apartado de registrar paciente, no se explica que algunos datos solo deben ser numéricos y otros solo texto.
P19	El reporte generado en Excel solo muestra una sesión realizado y no todo el histórico de sesiones del paciente, es confuso tener que generar un Excel cuando solo guardará un registro.

Tabla 5. Problemas de usabilidad identificados.

Una vez identificados los problemas, se procedió a realizar un análisis a los videos y notas de las pruebas que ofrecieran criterios para ser clasificados por el evaluador (facilitador de la prueba) como: problema menor (PM), problema medio (PMED), pérdida de funcionalidad (PF), fallo en la tarea (FT) y molestia (M), tal como se solicita y reporta en Lauesen (2005) (véase Tabla 6).

Problema	U1	U2	U3	Frecuencia	Clasificación Global
P1	PM			1	PM
P2	PM	PMED	PM	3	PM
P3	PF	PF	PF	3	PF
P4	PM	PM	PM	3	PM
P5	PM		PM	2	PM
P6	PM	PMED		2	PMED
P7	FT	FT	FT	3	FT
P8		PM		1	PM
P9	PM	PMED	PME D	3	PMED
P10		PM	PM	2	PM
P11			PM	1	PM
P12			PM	1	PM
P13			PM	1	PM
P14	PM			1	PM
P15			PM	1	PM
P16		PM		1	PM
P17			PM	1	PM
P18	PM	PMED	PM	3	PM
P19	PM		PM	2	PM

Table 6. Clasificación de los problemas de usabilidad, formato sugerido en Lauesen (2005).Fuente: Autoría propia.

En la Tabla 7 se presenta la relación entre las tareas y los problemas de usabilidad detectados. Durante la experiencia de la recolección de los problemas de usabilidad destacó la identificación del problema P2, que se relacionó con las tareas T7 y T8. Mediante estas tareas, los especialistas configuraban las características de la terapia, tales como, especificar el ángulo de flexión que el paciente debía realizar en cada repetición. Sin embargo, los terapeutas señalaron que el sistema no ofrecía ningún tipo de instrucción de cómo se debía realizar correctamente el ejercicio, es decir, el sistema necesitaba retroalimentar *feedback* al paciente sobre los movimientos (de hombro, de codo, de muñeca, entre otros) que debe involucrar para conseguir una repetición, ya que algunas tareas como el cepillado de dientes lo requieren.

Tarea	Problemas de usabilidad identificados
T1	P9, P11, P12, P18
T2	P5
T3	P3, P4
T4	P6, P7, P14, P15
T5	P2, P10, P19
T6	P1
T7	P2, P10, P17
T8	P2, P10, P17
T9	P6, P8, P13, P15
T10	P16

Tabla 7. Relación entre las tareas de usabilidad y los problemas identificados. Fuente: Autoría propia.

Por su parte, el problema P3, fue detectado durante la ejecución de la tarea T3, los terapeutas indicaron que a nivel semántico, el concepto de “eliminar un registro” no es correcto, ya que ellos como terapeutas no deben perder ningún registro sobre el seguimiento de algún paciente dado que en cualquier momento pueden regresar con ellos por alguna otra lesión y requerirían conocer sus antecedentes médicos, por lo que la expresión semántica viable sería, “dar de alta al paciente” indicando con ello que aun cuando a nivel del sistema se hubiese terminado el proceso de terapia, los registros sobre los mismos deberían seguir existiendo en el histórico pero señalando de alguna manera que el paciente ha sido dado de alta curado y por tanto no requiere terapia. Este hecho involucró también al problema P4, a lo cual los terapeutas sugirieron considerar dos apartados dentro del sistema, uno que involucre a los pacientes dados de alta y otro con los pacientes que se encuentran recibiendo terapia.

En el caso del problema P5, surgió a raíz de que los usuarios no pudieron consultar la información de un paciente debido a que dicha búsqueda se realizaba a partir de un código de identificación, el cual ellos tenían que buscar con apoyo de la barra de desplazamiento de la pantalla, para que una vez identificado el código lo anexarían en la pantalla de inicio de sesión para ingresar a la información del paciente, por lo que sugirieron realizar una búsqueda más efectiva por nombre y apellido del paciente que buscando uno a uno. Este problema no fue detectado de inmediato, hasta el momento de la participación oportuna del facilitador gracias a su Intervención Activa, quién les ofreció el escenario en el que se tuviesen 1000 pacientes registrados, hecho que los llevó a rectificar sobre la necesidad de mejorar el proceso de búsqueda.

Respecto al problema P7, este se activó durante la ejecución de la tarea T4, ya que los terapeutas requerían configurar la prueba del paciente ingresando con el código de éste último en el inicio de sesión mientras ellos lo buscaban dentro de su apartado como terapeutas, por lo que no se imaginaron que tenían que realizar una búsqueda por código de paciente, salir de su sesión y acceder a la sesión del paciente con dicha clave para realizar la respectiva configuración, lo que al final terminó clasificado como un fallo en la tarea, ya que esa actividad no pudo ser realizada por ninguno de los terapeutas.

Particularmente el problema P14, ofreció una impresión más allá de la funcionalidad del software, ya que estaba vinculada hacia el diseño accesible y basado en el usuario, es decir, el software fue diseñado para personas que no tuviera problemas visuales, hecho que representó un error, ya que el simple formato de letras y colores elegidos para el encabezado de las pantallas fue suficiente para ser detectado como un problema de usabilidad, ya que la Usuario 1 (U1) durante la tarea T4 se percató de su dificultad para leer los títulos de las pantallas que se encontraban en un fondo blanco y delineadas en negro, el terapeuta usa lentes y su problema visual combinado con la elección del formato seleccionado para el texto le impidió corroborar el sitio en el que se encontraba en una determinada pantalla (es decir, no estaba seguro de si se encontraba en la sesión del paciente porque el texto que se lo decía no era legible).

Esto evidentemente no se trata del modelo mental del usuario final, sino de un característico propio como perfil de usuario que no fue valorado durante el diseño del entorno y que para futuras mejoras debe ser tomado en cuenta.

A partir de los resultados presentados en la Tabla 6, se realizó un análisis estadístico mediante gráficos de barras. En la Figura 4 se encuentra representado el número de problemas de usabilidad identificados por los tres usuarios participantes junto con la frecuencia por tipo de problema en el que se clasificó; pudiendo notarse que el Usuario 3 (U3) fue el que identificó un mayor número de problemas de usabilidad con un total de 14, considerando que el total de los problemas detectados fueron 19, por lo que la variación de problemas de usabilidad detectados por U3 respecto a los otros dos usuarios solo dependen de los cinco problemas restantes, coincidiendo con el señalamiento de Nielsen (1993) sobre el número de usuarios necesarios para identificar tantos problemas de usabilidad como estos existan. Dentro del mismo gráfico se puede observar que la mayoría de los problemas detectados fueron clasificados como problema menor (PM), lo que indica que los usuarios pudieron concluir con la tarea de usabilidad asociada, pero en un tiempo mayor al esperado, señalando detalles de diseño que pueden corregirse en breve tiempo.

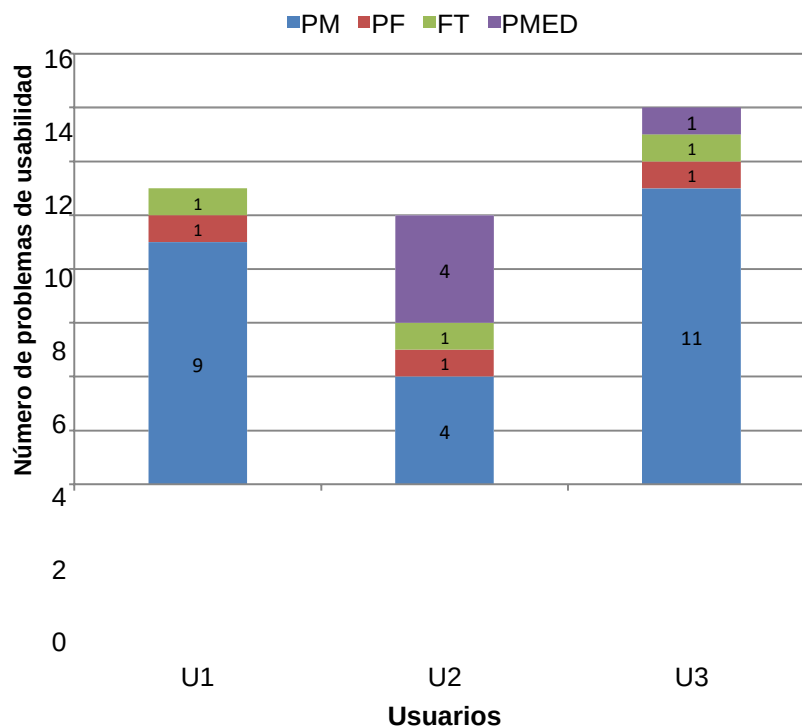


Figura 4. Frecuencia de problemas de usabilidad detectados por usuario. Fuente: Autoría propia.

Por su parte, en la Figura 5 se muestra la frecuencia de clasificación global dada a cada problema de usabilidad respecto a la moda de la clasificación en la que recayó una vez identificada por los diferentes usuarios finales, es decir, si un mismo problema de usabilidad fue identificado por los tres usuarios, la clasificación más frecuente es la que se consideró en la clasificación global ubicada en la última columna a la derecha de la Tabla 6. Este hecho confirma la presencia de un mayor número de problemas de usabilidad con clasificación Problema Menor (PM), lo cual también es visible en el Figura 4, de igual forma, se reafirma la presencia de un problema clasificado como pérdida de funcionalidad (PF) y otro como fallo en la tarea (FT), los cuales corresponden a los problemas P3 y P7 respectivamente, y que requieren especial atención y pronta corrección para obtener una mejor versión del software.

En la Figura 6 se representa la frecuencia de usuarios que pudieron detectar cada uno de los problemas de usabilidad, sobresaliendo entre ellos los problemas P2, P3, P4, P7, P9, y P18, los cuales fueron identificados por los tres usuarios y por ende serán problemas a priorizar, al ser reconocidos por su gravedad y/o su frecuencia de detección entre los usuarios.

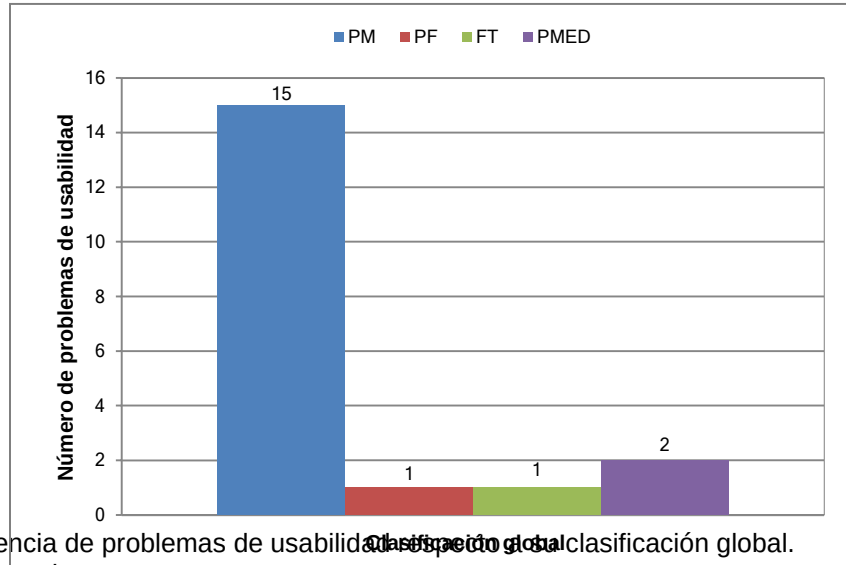


Figura 5. Frecuencia de problemas de usabilidad clasificados por clasificación global.
Fuente: Autoría propia.

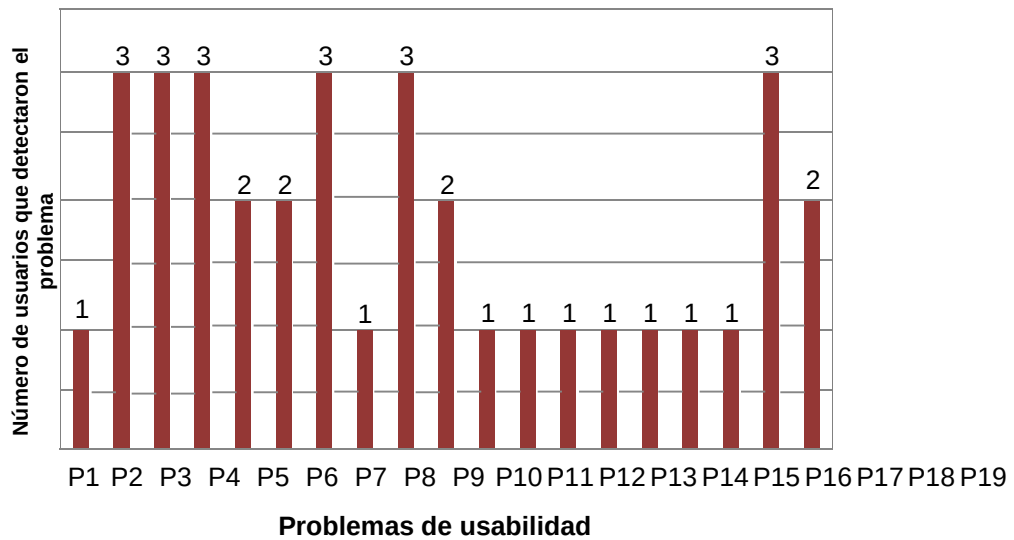


Figura 6. Frecuencia de usuarios que detectaron cada problema de usabilidad
Fuente: Autoría propia.

De la misma forma, en la Figura 7 se representa la frecuencia de problemas detectados por cada tarea, trascendiendo las tareas T1, T4, y T9 como las más informativas, con un total de cuatro problema de usabilidad cada una, lo que en conjunto representa más del 50% del total de los problemas de usabilidad detectados, lo que se corrobora en el despliegue de la Figura 8, marcando un precedente de los escenarios de evaluación que deben ser revisados con mayor detenimiento para evitar futuros problemas de usabilidad.

Por tanto, y como propuesta de solución para el conjunto de problemas de usabilidad detectados, en la Tabla 8, se presenta, el orden de prioridad bajo el cual se deberán atenderse los problemas y la solución a nivel de diseño de interfaz que las resolverá. Tomando en consideración que la mayoría de los problemas fueron clasificados con un nivel de gravedad leve, se considera que es posible resolver todas las necesidades especificadas por los usuarios finales.

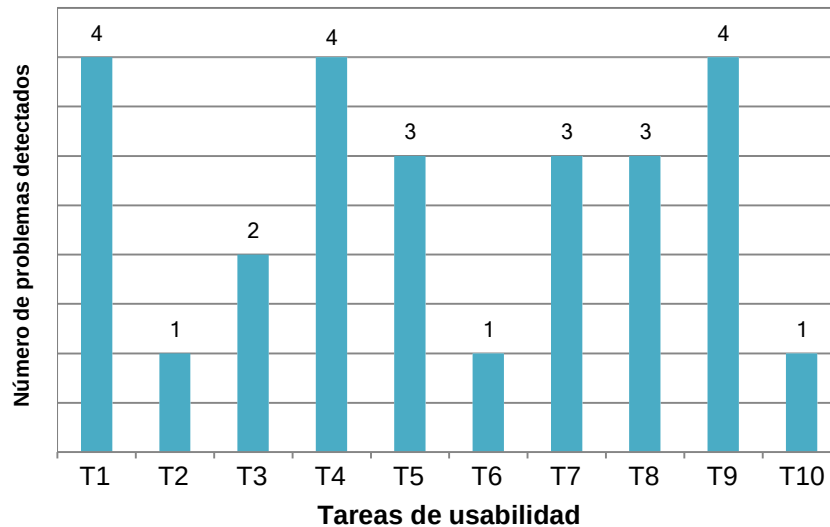


Figura 7. Frecuencia de problemas detectados por cada tarea de usabilidad
Fuente: Autoría propia.

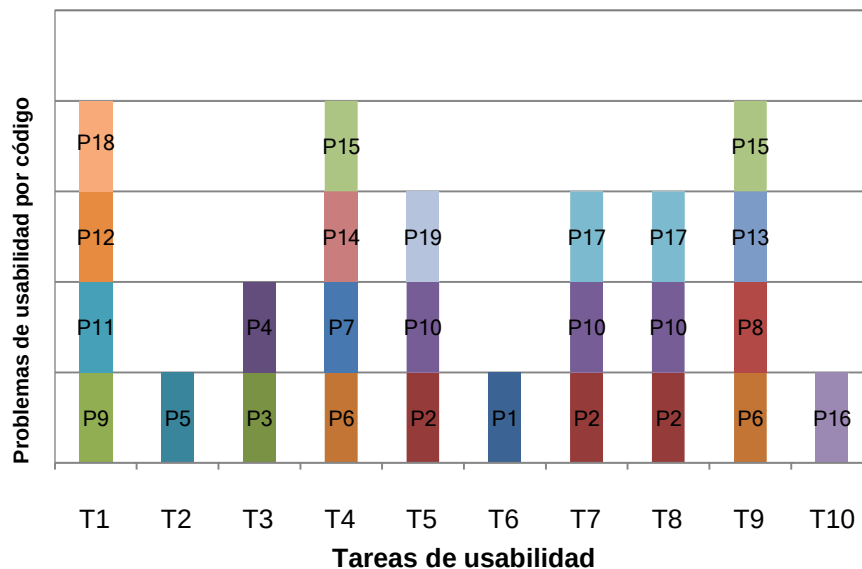


Figura 8. Frecuencia de problemas detectados por cada tarea de usabilidad rotulados por los problemas de usabilidad.
Fuente: Autoría propia.

6. Conclusiones

En este artículo se describió el plan de pruebas efectuado mediante el protocolo Pensando en Voz Alta en su modalidad de verbalización activa al módulo de terapeuta de un ambiente virtual no inmersivo que ofrece terapia basada en actividades de la vida diaria a pacientes con pérdida de movilidad en el codo debido a un ictus.

Como resultado de la prueba de usabilidad, se lograron recolectar 19 problemas de usabilidad, dentro de los cuales dos fueron considerados directamente de suma prioridad debido a que representaron un fallo en la tarea y pérdida de funcionalidad; cinco con prioridad media ya que fueron identificadas por la mayoría de los usuarios y doce de prioridad baja que están orientadas a pequeñas modificaciones a nivel de interfaz. Con ello, se logró un primer acercamiento a la expectativa del usuario potencial respecto al ambiente virtual bajo estudio, con miras a obtener una mejor versión en una próxima iteración del desarrollo del software, que permita reducir o anular los problemas de usabilidad identificados.

Prioridad	Razón	Problema de usabilidad	Solución
Alta	Al clasificarse como FT	P3	Dentro del apartado del terapeuta se modificará su interfaz para que pueda especificar al paciente que desea dar de alta con ayuda de un botón.
Alta	Al clasificarse por PF	P7	Una vez que encuentren al paciente en su lista de búsqueda se implementará un botón que le permita al terapeuta dirigirse al menú de acciones de su paciente y acceder prontamente a su histórico.
Media	Al clasificarse como PMED	P9	Con la frase “Tiempo en terapias” se almacenará la información concerniente a “cuánto tiempo estará en terapias el paciente” ya que consideran los terapeutas más importantes que la idea inicial. No hay cambio a nivel de la interfaz pero si en la definición del diccionario de datos de la base de datos que el software consume.
Media	Al clasificarse como PMED	P6	La función “volver” en la búsqueda de pacientes regresará a la pantalla previa y no a la pantalla de inicio de sesión, es un cambio a nivel de la interfaz.
Media	Al ser identificado por los 3 usuarios	P2	Se agregará a cada AVD un video explicativo de la manera en que se debe efectuar el ejercicio remarcando con colores los ángulos de flexión en las diferentes partes del brazo que estén involucradas.
Media	Al ser identificado por los 3 usuarios	P4	Desaparecerá de la interfaz de usuario la opción de eliminar paciente y se cambiará por la de darle de alta, con ello se impedirá la eliminación de registros.
Media	Al ser identificado por los 3 usuarios	P18	Se agregarán apartados de autoayuda para explicar los tipos de datos y restricciones de cada apartado en el formulario para registrar a un paciente.
Baja	Al ser identificado por los 2 usuarios	P5	Se modificará la interfaz del terapeuta para realizar la búsqueda del paciente por nombre y apellido.

Baja	Al ser identificado por los 2 usuarios	P19	Se trasladará el apartado de descarga a la sesión del usuario donde tenga la información de todas sus terapias para que se descargue en una hoja de cálculo todo el histórico si así se desea y se eliminará esa opción del apartado en el que el paciente terminaba su terapia.
Baja	Repercute en tres tareas de usabilidad	P10	Se integrarán instrucciones de uso en todas la interfaces de usuario para apoyar al usuario final en el proceso de gestión del software.
Baja	Repercute en dos tareas de usabilidad	P15	Pese a la existencia de un botón denominado “iniciar” que debe ser presionado para evaluar la clave introducida y así iniciar sesión, uno de los especialistas indicó la necesidad de poder ejecutar esa acción mediante la presión del botón <i>enter</i> , por lo que se programará dicha funcionalidad.
Baja	Repercute en dos tareas de usabilidad	P17	Se indagará en el código fuente de Unity con el Kinect para lograr agrandar el visor en espejo que contiene y pueda apreciarse mejor desde la aplicación (dicho visor beneficia al especialista para inspeccionar que el paciente esté realizando correctamente la terapia).
Baja	Afecta a una tarea involucrada con más problemas de usabilidad	P8	Se redefinirá a nivel de código de programación para que el histórico de citas de un paciente aparezca de la cita más reciente a la más antigua.
Baja	Afecta a una tarea involucrada con más problemas de usabilidad	P11	Se agregará al formulario de registro del paciente un cuadro de texto que permita detallar el evento que le produjo el ictus.
Baja	Afecta a una tarea involucrada con más problemas de usabilidad	P12	Se cambiará el <i>script</i> o <i>plugin</i> para el manejo del calendario.
Baja	Afecta a una tarea involucrada con más problemas de usabilidad	P13	Se implementará un menú de opciones en la sesión del administrador con la finalidad de que tenga acceso rápido a las diferentes tareas que puede realizar con su interfaz.
Baja	Afecta a una tarea involucrada con más problemas de usabilidad	P14	Se cambiarán los formatos difíciles de leer por colores y tipografía más claros.
Baja	Afecta a una tarea de usabilidad	P1	Se agregará mensaje de confirmación antes de poder salir del sistema.
Baja	Afecta a una tarea de usabilidad	P16	Se anejará a nivel de base de datos un nuevo campo que registre el nombre del último terapeuta que atendió al paciente y se actualizará dentro del formulario de configuración de la terapia del paciente, de tal manera que aparecerá en el histórico de terapias realizadas.

Tabla 8. Solución a problemas de usabilidad ordenados por prioridad. Fuente: Autoría propia.

Por su parte, el método de introspección denominado Pensando en Voz Alta logró cumplir con su propósito de entendimiento del modelo mental del usuario final (terapeuta), respecto al del programador-diseñador, haciendo notoria su rápida aplicación y bajo costo económico. Dicha estrategia permitió identificar los factores a nivel de diseño e inclusive de implementación que pueden promover el éxito de su implantación una vez liberado el software. A su vez, se logró tener testimonio de la crucial intervención del moderador durante la aplicación del protocolo, debido a que tuvo intervenciones oportunas que permitieron confirmar ciertos detalles de diseño que se sospechaban podrían causar problemas de usabilidad, como es el caso de la búsqueda del paciente a través de un identificador en vez de realizar la búsqueda por nombre y apellido, donde el moderador requirió ofrecer un escenario a futuro para que el usuario potencial confirmara la necesidad de hacer ajustes en el diseño. Los resultados obtenidos revelan que el módulo del terapeuta del ambiente virtual cubre las expectativas del usuario potencial, salvo la reestructuración necesaria a nivel de diseño y conceptual algunos elementos dentro de la interfaz, junto a la adhesión de algunas metáforas, lo que permitiría en una iteración subsecuente aplicar pruebas indagatorias o de *testing*, en conjunto con pruebas funcionales. Así, como trabajo a futuro se esperan aplicar al ambiente virtual los cambios identificados en la Tabla 8, bajo una nueva iteración del DCU, recabando resultados que permitan la liberación del software.

7. Agradecimientos

Agradecemos a la Clínica Therafix y al equipo de trabajo dirigido por el licenciado en terapia física Adrián Castillejos Pineda por su amable y entera disposición para la aplicación de las pruebas de usabilidad y con ello la recolección de datos que fomentaron la discusión de resultados presentados.

8. Referencias

- Acevedo, M. (2018). *Ambiente virtual para la rehabilitación pos-ictus de los movimientos del codo en adultos*. Universidad del Istmo, Santo Domingo Tehuantepec, México.
- Aguilar Acevedo, F., Pacheco Bautista, D., Acevedo Gómez, M., Toledo Toledo, G., & Nieva García, O. S. (2022). User-Centered Virtual Environment for Post-Stroke Motor Rehabilitation. *Journal of Medical Devices*. <https://doi.org/10.1115/1.4053605>
- Ávila García, M. S., Bianchetti, M., & González Gaviña, A. (2017). Uso del método "Thinking Aloud" en la investigación cualitativa. *Pistas Educativas*, 39(127), 26–38.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Bermudez i Badia, S., Fluett, G., Llorens, R., & Deutsch, J. (2016). Virtual reality for sensorimotor rehabilitation post stroke: Design principles and evidence. In D. Reinkensmeyer & V. Dietz (Eds.), *Neurorehabilitation Technology* (Second, pp. 573–603). Springer International Publishing. https://doi.org/https://doi.org/10.1007/978-3-319-28603-7_28
- Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3), 261–278. <https://doi.org/10.1109/47.867942>
- Dumas, J. S., & Redish, J. C. (1999). *A Practical Guide to Usability Testing* (1st ed.). Intellect Press.
- Hung, Y.-X., Huang, P.-C., Chen, K.-T., & Chu, W.-C. (2016). What Do Stroke Patients Look for in Game-Based Rehabilitation: A Survey Study. *Medicine*, 95(11), e3032. <https://doi.org/10.1097/MD.0000000000003032>
- Jääskeläinen, R. (2010). Think-aloud protocol. In Y. Gambier & L. Van Doorslaer (Eds.), *Handbook of Translation Studies* (pp. 371–373). John Benjamins Publishing Company.
- Korn, O., & Tietz, S. (2017). Strategies for Playful Design when Gamifying Rehabilitation: A Study on User Experience. *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*, 209–214. <https://doi.org/10.1145/3056540.3056550>
- Krahmer, E., & Ummelen, N. (2004). Thinking About Thinking Aloud: A Comparison of Two Verbal Protocols for Usability Testing. *IEEE Transactions on Professional Communication*, 47(2), 105–117. <https://doi.org/10.1109/TPC.2004.828205>
- Lauesen, S. (2005). *User interface design: a software engineering perspective*. Pearson Education.
- Nielsen, J. (1993). *Usability Engineering*. Morgan Kaufmann Publishers Inc.
- Nielsen, J. (2012). *Thinking Aloud: The #1 Usability Tool*. Nielsen Norman Group.

- <https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/>
- Nørgaard, M., & Hornbæk, K. (2006). What Do Usability Evaluators Do in Practice?: An Explorative Study of Think-aloud Testing. *Proceedings of the 6th Conference on Designing Interactive Systems*, 209–218. <https://doi.org/10.1145/1142405.1142439>
- Olmsted-Hawala, E. L., Murphy, E. D., Hawala, S., & Ashenfelter, K. T. (2010). Think-aloud Protocols: A Comparison of Three Think-aloud Protocols for Use in Testing Data-dissemination Web Sites for Usability. *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, 2381–2390. <https://doi.org/10.1145/1753326.1753685>
- Sáenz-de-Urturi, Z., García Zapirain, B., & Méndez Zorrilla, A. (2015). Elderly user experience to improve a Kinect-based game playability. *Behaviour & Information Technology*, 34(11), 1040–1051. <https://doi.org/10.1080/0144929X.2015.1077889>
- Shiratuiddin, M. F., Rai, S., Krishnan, G. M., Newton, M., Wang, X., Soheli, F., Blacker, D., & Byrnes, M. (2017). A usability evaluation of Neuromender's upper limb game-based rehabilitation system for stroke survivors. In N. Dias (Ed.), *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)* (pp. 1–8). IEEE. <https://doi.org/10.1109/SeGAH.2017.7939258>
- Someren, M. van, Barnard, Y., & Sandberg, J. (1994). *The Think Aloud Method - A Practical Guide to Modelling Cognitive Processes*. Academic Press.
- Toledo, T. G., Rosales, L. R., Hernandez, M. A., & Arellano, J. P. J. (2015). Usability test of a SCADA system for monitoring and operating an experimental wind rotor. *2015 International Conference on Computing Systems and Telematics, ICCSAT 2015*. <https://doi.org/10.1109/ICCSAT.2015.7362932>
- Van Someren, M. W., Barnard, Y. F., y Sandberg, J. A. C. (1994). *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*. London: Academic Press.
- Yoshida, K., Asakawa, K., Yamauchi, T., Sakuraba, S., Sawamura, D., Murakami, Y., y Sakai, S. (2013). The Flow State Scale for Occupational Tasks: Development, Reliability, and Validity. *Hong Kong Journal of Occupational Therapy*, 23(2), 54–61. <https://doi.org/https://doi.org/10.1016/j.hkjot.2013.09.002>



Recibido 14 Mar. 2023

ReCIBE, Año 12 No. 1, May. 2023

Aceptado 03 Abr. 2023

La presencia docente como determinante de la presencia cognitiva en un curso Blended Learning basado en la Comunidad de Indagación

Teaching presence as a determinant of cognitive presence in a Blended Learning course based on the Community of Inquiry

Julio Erick Medina López^{1*}
jemedina@docentes.uat.edu.mx

Héctor Martínez Ruiz²
hector.martinezr@uaq.mx

Ezra Federico Parra González³
ezra.parra@cimat.mx

Jezreel Mejía Miranda³
jmejia@cimat.mx

Universidad Autónoma de Tamaulipas¹
Universidad Autónoma de Querétaro²
Centro de Investigación en Matemáticas A. C. Unidad Zacatecas.³
Contacto principal. *

Resumen

En el presente trabajo se presentan los resultados del análisis de la influencia de la presencia docente sobre la presencia cognitiva en un curso de Blended Learning. Se utilizó el modelo de Comunidad de Indagación como referencia, el cual es un enfoque constructivista para el diseño de experiencias educativas que integra tres tipos de presencia: presencia docente, presencia social y presencia cognitiva. Se diseñó un curso en formato Blended Learning para una asignatura en la Universidad Autónoma de Tamaulipas. La metodología empleada incluyó la validación y aplicación de la escala de Comunidad de Indagación para obtener la percepción de las tres presencias en el diseño del curso, y se realizó un análisis correlacional para determinar la influencia que tuvo la presencia docente (diseño instruccional, facilitación del discurso e instrucción directa) en la percepción de la presencia cognitiva. Los resultados muestran una fuerte relación entre la presencia docente y la presencia cognitiva ($r=0.771$, $p=0.000$), lo que indica una correlación significativa. Además, se encontró una relación significativa entre los elementos de la presencia docente y la presencia cognitiva. Estos hallazgos confirman la importancia de la planificación cuidadosa y anticipada del curso es fundamental para garantizar el éxito de la Comunidad de Indagación en el contexto del Blended Learning en la educación superior.

Palabras clave: *Blended Learning, Presencia docente, Presencia cognitiva, Comunidad de Indagación.*

Abstract

This paper presents the results of the analysis of the influence of teaching presence on cognitive presence in a Blended Learning course. The Community of Inquiry model was used as a reference, which is a constructivist approach for the design of educational experiences that integrates three types of presence: teaching presence, social presence and cognitive presence. A Blended Learning course was designed for a subject at the Universidad Autónoma de Tamaulipas. The methodology employed included the validation and application of the Community of Inquiry scale to obtain the perception of the three presences in the course design, and a correlational analysis was performed to determine the influence of teaching presence (instructional design, discourse facilitation and direct instruction) on the perception of cognitive presence. The results show a strong relationship between teaching presence and cognitive presence ($r=0.771$, $p=0.000$), indicating a significant correlation. In addition, a significant relationship was found between elements of teaching presence and cognitive presence. These findings confirm the importance of careful and advance planning of the course is critical to ensure the success of the Community of Inquiry in the context of Blended Learning in higher education.

Keywords: *Blended Learning, Teaching Presence, Cognitive presence Community of Inquiry.*

1. Introducción

Desde el punto de vista de Wiley (2006), el mundo se encuentra en constante cambio y considera que la economía y la ciencia se han adaptado a través del movimiento de lo analógico a lo digital, de lo cerrado a lo abierto, de lo fijo a lo móvil, de lo genérico a lo personalizado y del consumo a la participación. Sin embargo, en lo referente a la educación superior, estos argumentos también son válidos en el sentido que la educación superior debe ir a la par de movimientos que implican cambios pertinentes y que se ajusten al tipo de estudiantes que se forman en las Universidades y los efectos que sobre ellos tienen las transformaciones de la sociedad actual y también invita a pensar si las modalidades tradicionales, la educación a distancia, el e-learning o el aprendizaje en línea o presencial se adaptan a las condiciones actuales, puesto que aquellos estudiantes nacidos y formados con el uso cotidiano de la tecnología digital y que se han denominado “nativos digitales” consideran que han llegado a las aulas para instruirlos, “desconocidos” que hablan otros idiomas y se denominan “inmigrantes digitales” (Prensky, 2001), por lo que se genera una especie de brecha generacional que implica enseñar y aprender de forma diferente. Así, a partir del análisis de los elementos que ofrece cada modalidad, de sus ventajas y desventajas, es posible obtener un panorama más amplio de lo que podría integrarse de cada modelo en una nueva propuesta como alternativa para solventar las necesidades en un contexto educativo en particular. Por ejemplo, en la enseñanza tradicional, se propicia el aprendizaje cuando se integran diversos recursos, entre estos la tecnología, para generar nuevas estrategias, que promuevan el abandono de prácticas tradicionales descontextualizadas que, aún y con la disposición de la tecnología, siguen prevaleciendo, con la tradición de cumplir con instrucciones para la realización de tareas (Martínez et al., 2018).

Es por lo anterior que el profesor como un *actor que construye los ambientes de aprendizaje* debe adquirir ciertas capacidades que contribuyan favorablemente en la interacción, sean estos ambientes físicos o virtuales. Entre esas características de acuerdo con (Mendoza and Abaunza, 2005), se pueden mencionar: *la tolerancia, el deseo de aprender constantemente, el liderazgo, la honestidad, la sencillez y la disposición.*

Por tanto, la *capacidad motivadora* del docente y el éxito de los esfuerzos por lograr un aprendizaje podría depender también de estas y otras cualidades como la capacidad comunicativa y la empatía, dado que estos elementos pueden ser parte de la *mediación pedagógica* entendida la forma que se tratan los contenidos para hacer posible el acto educativo (Mendoza and Abaunza, 2005).

Así, la educación tradicional por lo regular ocurre en el mismo lugar y en el mismo momento o en diferente lugar en diferente momento principalmente desarrollándose en un espacio físico y autónomo que suele estar centrado en el docente y que tiene sus ventajas, igual que la educación a distancia, sea ésta geográfica, temporal o intelectual (Simonson et al., 2015), por lo se concluye que un modelo no es mejor que el otro, sino que depende de la calidad y de los fines (Romero-Mayoral et al., 2014), así como la capacidad del docente para construir un ambiente que favorezca el aprendizaje al llevar a cabo un diseño instruccional adecuado, fomentar el diálogo, la participación y la motivación así como la instrucción.

Por otra parte, la educación a distancia la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), define la educación a distancia como “un proceso y un sistema educativo en el que toda o una parte importante de la enseñanza es impartida por alguien o algo alejado en el espacio y el tiempo del alumno” (UNESCO, n.d.), y que requiere una planificación estructurada, cursos bien diseñados, técnicas instruccionales especiales y métodos de comunicación electrónicos y de otro tipo cuya objetivo es incluir diferentes dimensiones de apertura y flexibilidad en términos de acceso, currículo y otros elementos estructurales (Burns, 2011).

Por su parte, Fillip (2001), define a la educación a distancia como “un proceso educativo en el que una parte importante parte de la enseñanza es impartida por alguien alejado en lugar y/o tiempo, del el alumno”, lo que implica que las formas en las que se desarrolla son diversas, entre la que señala la periodicidad de las sesiones, el nivel de apoyo del instructor o tutor, la tecnología, la institución y la naturaleza del curso. De esta manera, los fundamentos de la educación a distancia son el soporte para diferentes metodologías entre las que se encuentra el Blended Learning, entre estos fundamentos que se encuentran el otorgar mayor responsabilidad al estudiante sobre su aprendizaje (Barberà et al., 2006) y centrar al profesor en tareas exclusivamente educativas.

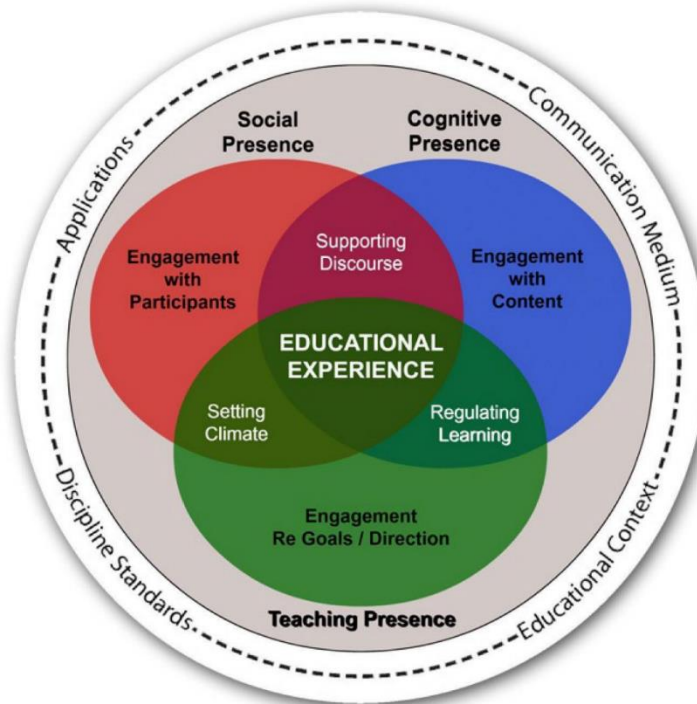
Los cursos en modalidad Blended Learning también denominados mixtos o combinados, permiten integrar elementos específicos de la enseñanza presencial, virtual y a distancia para crear una nueva configuración del ambiente de aprendizaje que resulte efectivo, eficiente y flexible para sus alumnos (Stein and Graham, 2014) con amplio apoyo de las Tecnologías de la Información Comunicación (Meydanlioglu and Arikan, 2014), de tal manera que los estudiantes, los docentes y el contenido, interactúen de forma ilimitada en espacio y tiempo (Adini et al., 2017).

Hay diversas afirmaciones sobre las formas que adoptan los modelos Blended Learning. De acuerdo con Hofman (2018), esta modalidad se puede representar como una serie de bloques secuenciados para crear experiencias de aprendizaje a través de un plan que da seguimiento de principio a fin ajustando los objetivos educativos al ambiente de aprendizaje más adecuado. De cualquier forma, es el docente que establece su presencia de diferentes maneras, principalmente como constructor del ambiente o experiencia de aprendizaje, como facilitador y como guía hacia la comprensión y el aprendizaje significativo.

La diferentes configuraciones dan lugar a un repertorio de modelos de implementación del Blended Learning, entre los que se encuentran: el modelo de bloques (Cleveland. and Wilton, 2018) , modelo de flujo de programa, núcleo radios (Bersin, 2004), modelo de presentación interacción (Cleveland. and Wilton, 2018), modelo de rotación, flex y a la carta (Horn y Staker, 2015) y modelos híbrido flexibles y modelo de Comunidad de Indagación.

El modelo de Comunidad de Indagación es un modelo de proceso colaborativo-constructivista que describe los elementos esenciales para el éxito de una experiencia de aprendizaje en línea en la enseñanza superior, basado en la filosofía educativa de Dewey y en el constructivismo social (Castellanos-Reyes, 2020). En este modelo, se establecen tres presencias para el desarrollo de una experiencia de aprendizaje: presencia de la enseñanza, presencia social y presencia cognitiva.

Figura 1 Modelo de Comunidad de Indagación (Col).



Fuente: Garrison et al. (2000).

La presencia docente tiene que ver con el papel que asume el profesor al diseñar, facilitar y dirigir los procesos cognitivos y sociales para el aprendizaje entre pares, apoyo al aprendizaje y retroalimentación. La presencia cognitiva se relaciona con la forma en que se genera un pensamiento crítico a través del discurso y reflexión y la presencia social promueve al trabajo colaborativo en línea y se define como la capacidad de los estudiantes de sentirse afectivamente conectados con sus compañeros y presentarse con una personalidad propia a través de la comunicación mediada por tecnología (Castellanos-Reyes, 2020).

La presencia docente en el modelo de Comunidad de Indagación se compone de tres elementos: (1) el diseño instruccional y organización (por ejemplo, establecer el programa, diseñar estrategias, etc.); (2) facilitar el discurso (por ejemplo, establecer el clima, reconocer y reforzar las contribuciones de los estudiantes, etc.) y (3) la instrucción directa (resumir la discusión, presentar preguntas o contenidos, etc.) (Anderson et al., 2001) citados por (Fiock, 2020).

La presencia cognitiva la integran los elementos del Modelo de Indagación Práctica (PIM, por sus siglas en inglés) que plantea el Col y que corresponden a un evento desencadenante, exploración, integración y resolución, elementos que se desarrollan a través de la integración del mundo público y privado como elementos centrales para el desarrollo de la presencia cognitiva a través de las actividades de aprendizaje, que en esa secuencia, se presentan inicialmente como dilemas o problemas con los que los estudiantes se identifican para posteriormente buscar explicaciones mediante actividades grupales, lluvia de ideas o revisión de literatura, seguido de la reflexión más detenida para lograr la comprensión y finalmente plantear una posible solución de la tarea, lo que probablemente inicia nuevamente el ciclo de indagación (Garrison, 2017).

De acuerdo con Fiock, (2020), el papel del instructor es crear una historia narrativa o un camino a través del diseño del curso y el contenido del curso cuando se utiliza el Col como base para un diseño eficaz que promueva el logro de los aprendizajes y diversos estudios han determinado que la presencia docente es una responsabilidad y una condición que determina la presencia cognitiva (Garrison et al., 2010) referida como aquella que implica un aprendizaje profundo o habilidades de pensamiento de alto nivel (Mohammad et al., 2023). Además, se sugiere que el desarrollo del curso debe orientarse a la reflexión permanente en lugar de que el contenido sea absorbido de forma pasiva (Garrison, 2017).

2. Metodología

El diseño de la investigación fue transeccional correlacional. En este tipo de diseño, se describen las relaciones entre dos o más variables en un momento determinado y esta identificación de la correlación entre variables se realiza mediante la aplicación de métodos estadísticos. Para este caso se utilizó un método no paramétrico de correlación de Spearman al constituirse la muestra fuera de los parámetros de normalidad, al ser pequeña y cuyas observaciones se midieron de forma ordinal.

El coeficiente de correlación de Spearman es una asociación lineal que permite conocer el grado de asociación entre dos variables y además determinar si son interdependientes. Los valores obtenidos en el coeficiente cercanos a 1 indican que existe una fuerte asociación entre las variables o clasificaciones, los valores cercanos a -1 indican una fuerte asociación negativa y cuando el valor es 0 se establece que no hay relación (Montes et al., 2021). Además de lo anterior, se considera el valor de significancia **p** que cuando es menor a 0.05 indica una correlación significativa.

El objetivo del trabajo fue determinar si hay una relación entre la presencia docente y la presencia cognitiva en un curso bajo el modelo Blended Learning construido bajo los preceptos del modelo de Comunidad de Indagación.

Hipótesis

H0: La presencia docente contribuye de forma significativa a la presencia cognitiva en el Blended Learning bajo el modelo de Comunidad de Indagación.

H1: La presencia docente no contribuye de forma significativa a la presencia cognitiva en el Blended Learning bajo el modelo de Comunidad de Indagación.

Los participantes fueron estudiantes de cuarto periodo de la Licenciatura en Ciencias de la Educación, seleccionados de forma no probabilística de forma intencionada por el criterio de acceso. Debido a las condiciones pandémicas por COVID-19, el curso se llevó a cabo en modalidad híbrida con sesiones virtuales presenciales y trabajo asincrónico durante el primer periodo del año 2022 en la Universidad Autónoma de Tamaulipas.

3. Procedimiento

El periodo 2022-1, corresponde al primer periodo lectivo en la Universidad Autónoma de Tamaulipas. El trabajo académico implicó que el Blended Learning se realizara de forma virtual presencial y con trabajo asíncrono debido a que prevalecieron las restricciones por la pandemia del COVID. La presencia docente se estableció a través del *diseño y organización del curso*, la operacionalización de los ítems se llevó a cabo mediante descripción escrita de los objetivos instruccionales en cada unidad didáctica, presentación de los contenidos temáticos, las instrucciones en cada actividad de aprendizaje y el establecimiento de plazos de forma clara, utilizando la plataforma LMS y de forma verbal.

La *facilitación* se realizó mediante la interacción directa con los estudiantes a través de las sesiones virtuales presenciales, mediante la comunicación asincrónica y resolución de dudas, con las actividades de aprendizaje para indagar conceptos nuevos, así como la generación de un clima de confianza para atención personalizada.

La *instrucción directa* consideró el establecimiento de mecanismos de retroalimentación puntual al estudiante, tanto de forma verbal y escrita y por distintos medios, presenciales, virtuales sincrónicos y asincrónicos. Las herramientas de comunicación que se utilizaron fueron los sistemas de mensajería de las plataformas Moodle y Microsoft Teams, y la sesión sincrónica para comunicación en tiempo real. La retroalimentación se realizó para cada actividad de aprendizaje de forma inmediata, de tal forma que el estudiante reconociera sus avances. También se resolvieron dudas de forma grupal y en subgrupos cuando lo solicitaron fuera de clase.

En el caso de la presencia cognitiva, las actividades de aprendizaje se diseñaron de acuerdo con el modelo de indagación práctica del Col en las que se promovió el aprendizaje significativo a través de la discusión de conocimientos previos, preguntas detonadoras, búsqueda de nueva información y resolución de actividades de aprendizaje con elementos reflexivos al final de la actividad mediante la publicación en foros virtuales de la plataforma tecnológica de aprendizaje.

4. Instrumentos

Una vez terminada la intervención, para la recolección de los datos se utilizó la escala de Comunidad de Indagación. El instrumento contiene 34 ítems que valoran las tres presencias del modelo Col en una escala de cinco puntos Likert. Para la presencia de la enseñanza los ítems correspondientes son del 1 al 13 y para la presencia cognitiva los ítems son del 23 al 34. El instrumento fue validado previamente (Medina et al., 2022). La aplicación del instrumento permitió obtener los datos descriptivos para cada ítem y cada presencia. La encuesta se aplicó vía electrónica mediante Google Forms.

La encuesta se aplicó bajo un consentimiento informado sobre las implicaciones del estudio indicando, además, que la participación era voluntaria. La administración de la escala se llevó a cabo por el docente responsable del grupo y no se registraron datos que identifiquen a los participantes.

5. Resultados

La presencia docente se valoró con 13 (1-13) ítems de la escala Col y la presencia cognitiva con 12 (23-34), organizados en tres componentes: Diseño y organización (ítems 1-4), facilitación (ítems 5-10) e instrucción directa (ítems 11-13). Los estadísticos descriptivos de la Tabla 1, muestran que hay altos promedios en ambas presencias siendo la presencia docente la que tiene el promedio más alto posterior a la intervención.

Estadísticos descriptivos					
	N	Mínimo	Máximo	Media	Desviación
Presencia docente	18	3.46	5.00	4.6239	.52497
Presencia cognitiva	18	3.38	5.00	4.5046	.55353
N válido (por lista)	18				

Tabla 1 Estadísticos descriptivos de la Presencia docente y cognitiva.

En su conjunto, se observa que la presencia docente obtuvo una percepción tendiente al nivel superior de la escala (Totalmente de acuerdo) en la mayoría de los ítems, como se observa en la Tabla 2.

Tabla 2 Estadísticos descriptivos de todos los ítems de la presencia docente.

	Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo	Promedio
1.- El profesor comunicó con claridad los contenidos del curso.	66.7	33.3	0	0	0	4.67
2.- El profesor comunicó con claridad los objetivos del curso	77.8	22.2	0	0	0	4.78
3.- El profesor dio instrucciones claras sobre como participar en las actividades de aprendizaje del curso.	66.7	27.8	5.6	0	0	4.61
4.- El profesor dio a conocer claramente las fechas y plazos de entrega de las actividades de aprendizaje.	61.1	27.8	5.6	5.6	0	4.44
5.- El profesor me ayudó a identificar aspectos polémicos sobre los temas, lo que me ha ayudado a aprender	77.8	16.7	5.6	0	0	4.72
6.- El profesor tuvo la disposición para guiar la sesión hacia el entendimiento de los temas, lo que ayudó clarificar mi pensamiento	66.7	27.8	5.6	0	0	4.61
7.- El profesor ayudó a mantener a los estudiantes involucrados y participativos en un diálogo productivo	72.2	22.2	5.6	0	0	4.67
8.- El profesor ayudó a mantener a los participantes enfocados de tal manera que facilitó el aprendizaje	72.2	22.2	5.6	0	0	4.67
9.- El profesor alentó a los estudiantes a explorar nuevos conceptos del curso	66.7	33.3	0	0	0	4.67
10.- Las acciones del profesor reforzaron el desarrollo de un sentido de comunidad entre los participantes	72.2	22.2	5.6	0	0	4.67
11.- El profesor ayudó a centrar la discusión en temas relevantes de manera que me ayudó a aprender	72.2	22.2	5.6	0	0	4.67
12.- El profesor me proporcionó retroalimentación que me ayudó a entender mis puntos fuertes y débiles en relación con las metas y objetivos del curso	66.7	16.7	16.7	0	0	4.50
13.- El profesor proporcionó retroalimentación de manera oportuna	61.1	22.2	16.7	0	0	4.44

Para determinar la relación entre la presencia docente y la presencia cognitiva se obtuvieron las correlaciones entre las dos presencias que se muestran en la Tabla 3.

Tabla 3 Correlaciones entre la presencia docente y cognitiva

Correlaciones			Presencia docente	Presencia cognitiva
Rho de Spearman	Presencia docente	Coeficiente de correlación	1.000	.771**
		Sig. (bilateral)	.	0.000
		N	18	18
	Presencia cognitiva	Coeficiente de correlación	.771**	1.000
		Sig. (bilateral)	0.000	.
		N	18	18

****.** La correlación es significativa en el nivel 0,01 (bilateral).

De lo anterior, se identifica existe una alta correlación entre la presencia docente y la presencia cognitiva (0.771) lo que significa que la presencia cognitiva está fuertemente determinada por la presencia docente, que como se mencionó, se operacionaliza a través del diseño y organización del curso, facilitación e instrucción directa.

Al determinar cuál es la relación entre los componentes de la presencia docente y la correlación con la presencia cognitiva se obtuvieron las siguientes correlaciones (Tabla 4):

Tabla 4 Correlaciones entre la presencia cognitiva y los elementos de la presencia docente.

Correlaciones			Presencia cognitiva	Diseño y organización del curso	Facilitación	Instrucción directa
Rho de Spearman	Presencia cognitiva	Coeficiente de correlación	1.000	.750**	.774**	.737**
		Sig. (bilateral)	.	.000	.000	.000
		N	18	18	18	18
	Diseño y organización del curso	Coeficiente de correlación	.750**	1.000	.870**	.977**
		Sig. (bilateral)	.000	.	.000	.000
		N	18	18	18	18
	Facilitación	Coeficiente de correlación	.774**	.870**	1.000	.886**
		Sig. (bilateral)	.000	.000	.	.000
		N	18	18	18	18
	Instrucción directa	Coeficiente de correlación	.737**	.977**	.886**	1.000
		Sig. (bilateral)	.000	.000	.000	.
		N	18	18	18	18

****.** La correlación es significativa en el nivel 0,01 (bilateral).

Como se observa, todos los componentes de la presencia docente están relacionados de forma significativa con la presencia cognitiva. La presencia cognitiva está altamente relacionada con el diseño y organización del curso ($r=0.750$), la facilitación ($r=0.774$) y la instrucción directa ($r=0.737$). No obstante, el valor más alto de correlación entre la presencia cognitiva y los componentes de la presencia docente se encuentra en el componente de facilitación ($r=0.774$), por lo que deduce que los elementos relacionados con el rol del docente respecto a la guía, generación de diálogo y motivación fueron efectivas para determinar altos niveles de percepción positiva respecto a la presencia cognitiva.

6. Discusión y conclusiones

Los resultados comprueban la hipótesis de que la presencia docente es fundamental para predecir la presencia cognitiva y coinciden con los de (Dempsey and Zhang, 2019; Joo et al., 2011) al determinar que la presencia docente juega un rol significativo en la presencia cognitiva.

Las estadísticas descriptivas, denotan una percepción altamente positiva respecto a la presencia docente, sobre todo en aspectos relacionados con la claridad en las instrucciones y los tiempos de entrega, lo que coincide con los resultados de (Rao et al., 2015) quienes encontraron que el esfuerzo y tiempo destinado a la planificación es determinante para una percepción positiva de los estudiantes quienes enfatizaron la claridad y buena organización del curso, lo que repercute en la presencia cognitiva dado que el diseño instruccional para Blended Learning de establecerse a priori para garantizar que todas las actividades estén previamente planificadas antes de que inicie el curso.

De acuerdo con Gagné et al. (1992), en el desarrollo de una lección, existen una serie de eventos en los que se requiere la participación del docente en mayor o menor grado y es lo que se conoce como instrucción. Estos eventos establecen la comunicación con el estudiante. Esta comunicación en un ambiente Blended Learning puede ser verbal o escrita y el objetivo principal de ésta es apoyar los procesos internos de aprendizaje, por lo tanto, el rol del docente es determinante.

El primero de estos eventos está relacionado con la capacidad del profesor de ganar la atención de los estudiantes (Gagné et al., 1992). Se observa en los resultados que los estudiantes percibieron que las acciones del profesor los mantuvieron involucrados y participativos y que propició que se mantuvieran enfocados en las actividades lo que facilitó el aprendizaje. Esto coincide con los resultados de Zainol et al. (2018) y Toboso et al. (2017) que identifican que la utilización por parte del docente de una variedad de recursos necesarios para involucrar a los alumnos, es de suma importancia para los estudiantes y además tiene implicaciones en la presencia cognitiva como se observa en las correlaciones de cada elemento de la presencia docente con la presencia cognitiva.

Otro de los eventos en el proceso instruccional tiene que ver con guiar a los estudiantes hacia el aprendizaje, es decir la forma en que se estimula al estudiante hacia la comprensión. Al respecto se observó que perciben de forma favorable el apoyo del docente para llevar a los estudiantes hacia el entendimiento de los temas. Esto es particularmente relevante puesto que el grado de insinuación o estímulo que se emplea en el aprendizaje varía de un estudiante a otro (Gagné et al., 1992), ya que algunos necesitan más apoyo que otros en el proceso.

En conclusión, para la implementación del Blended Learning la presencia docente tuvo un papel determinante en cómo perciben el curso los estudiantes respecto a los elementos de la presencia cognitiva, puesto que la presencia docente involucra el diseño perottam paso en la configuración del ambiente de aprendizaje de una forma sistemática y racional para logro de los objetivos de aprendizaje.

Finalmente, se confirma que la presencia de la enseñanza materializada en el diseño instruccional, la facilitación y la instrucción directa son elementos imprescindibles y determinantes para establecer una Comunidad de Indagación y lograr que los estudiantes desarrollen el pensamiento crítico y un aprendizaje profundo con el apoyo del docente, quien, en su rol de facilitador o guía, establece los mecanismos para que se logren estas condiciones.

7. Referencias

- Adini, M. H., Purba, H. S. and Sukmawati, R. A. (2017). The Development of Blended Learning Model Using Wordpress. *5th SEA-DR (South East Asia Development Research) International Conference 2017 (SEADRIC 2017)*. <https://doi.org/10.2991/seadric-17.2017.45>
- Barberà, E., Romiszowski, A., Sangrà, A. and Simonson, M. (2006). *Educación Abierta y a Distancia*.
- Bersin, J. (2004). *The Blended Learning Book Best*. 351.
- Burns, M. (2011). *Distance Education for Teacher Training: Modes, Models, and Methods*. 338. [http://idd.edc.org/sites/idd.edc.org/files/Distance Education for Teacher Training by Mary Burns EDC.pdf](http://idd.edc.org/sites/idd.edc.org/files/Distance%20Education%20for%20Teacher%20Training%20by%20Mary%20Burns%20EDC.pdf)
- Castellanos-Reyes, D. (2020). 20 Years of the Community of Inquiry Framework. *TechTrends*, 64(4), 557–560. <https://doi.org/10.1007/s11528-020-00491-7>
- Cleveland, M. and Wilton, D. (2018). *Guide to Blended Learning*.
- Dempsey, P. R. and Zhang, J. (2019). Re-examining the construct validity and causal relationships of teaching, cognitive, and social presence in community of inquiry framework. *Online Learning Journal*, 23(1), 62–79. <https://doi.org/10.24059/olj.v23i1.1419>
- Fillip, B. (2001). *Distance Education in Central America and the Caribbean*. <https://www.uts.utoronto.ca/~chan/istb01/readings/distanceEdpdf>
- Fiock, H. S. (2020). View of Designing a Community of Inquiry in Online Courses | The International Review of Research in Open and Distributed Learning. *International Review of Research in Open and Distributed Learning*, 21(1), 1–19.
- Gagné, R., Briggs, L. and Wager, W. (1992). *Principles of Instructional Design*.
- Garrison, D. R., Anderson, T. and Archer, W. (2000). *Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education*.
- Garrison, D. R., Anderson, T. and Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *Internet and Higher Education*, 13(1–2), 5–9. <https://doi.org/10.1016/j.iheduc.2009.10.003>
- Garrison, R. (2017). *E-Learning in the 21st Century*.
- Hofman, J. (2018). *What works in talent development: Blended Learning*.
- Horn, M. B. and Staker, H. (2015). *Blended: Using Disruptive Innovation to Improve Schools* (Wiley (ed.)).
- Joo, Y. J., Lim, K. Y. and Kim, E. K. (2011). Online university students' satisfaction and persistence: Examining perceived level of presence, usefulness and ease of use as predictors in a structural model. *Computers and Education*, 57(2), 1654–1664. <https://doi.org/10.1016/j.compedu.2011.02.008>
- Medina, J. E., Martínez, H. and Sánchez, L. I. (2022). Validación de la escala de Comunidad de Indagación para la evaluación del Blended Learning. *Transdigital*, 3(6), 1–24. <https://doi.org/10.56162/transdigital131>
- Mendoza, F. J. and Abaunza, O. M. (2005). La mediación pedagógica: una nueva perspectiva en la formación de valores educativos. *Humanidades: La Ética En El Inicio Del Siglo XXI*, 1–11.
- Meydanlioglu, A. and Arikan, F. (2014). Effect of Hybrid Learning in Higher Education. *International Journal of Information and Communication Technology Education*, 8(2), 5. <https://doi.org/10.4018/jicte.2007040101>
- Mohammad, A., Al-iessa, A. H. K. A. and Afrouz, M. (2023). *Relationship between Iraqi EFL Students' Perceptions of Community of Inquiry and their Attention and Engagement in Online Classes*. 17(1), 227–255. <http://tel.journal.org>

- Montes, A., Ochoa, J., Juaréz, B., Vasquez, M. and Díaz, C. (2021). Aplicación del coeficiente de correlación de Spearman en un estudio de fisioterapia. *Cuerpo Académico de Probabilidad y Estadística BUAP*, 1–4. <https://www.fcfm.buap.mx/SIEP/2021/Extensos Carteles/Extenso Juliana.pdf>
- Prensky, M. (2001). Digital Natives, Digital Immigrants. *On The Horizon*, 9. <https://www.marcprensky.com/writing/Prensky - Digital Natives, Digital Immigrants - Part1.pdf>
- Rao, K., Edelen, P. and Cat, W. (2015). Universal design for online courses: applying principles to pedagogy. *Open Learning*, 30(1), 35–52. <https://doi.org/10.1080/02680513.2014.991300>
- Romero, J., Roca, C., Sanjuán, A. and Pulido, A. (2014). DISEÑO DE UN APRENDIZAJE ADAPTADO A LAS NECESIDADES DEL ALUMNO 1 Melchor GARCÍA-DOMÍNGUEZ. *Tesi*, 15(3), 1–255.
- Simonson, M., Smaldino, S., Albright, M. and Zvacek, S. (2015). *Teaching and Learning at a Distance (Sixth Edition)*. <http://learning.fon.edu.mk/knigi/teachinganlearningatadistance-4.pdf>
- Stein, J. and Graham, C. R. (2014). Orientation to Blended Teaching and Learning. In Routledge (Ed.), *Essentials for Blended Learning* (p. 22).
- Toboso, S., Tello, I. and Alvarez, F. J. (2017). Formación musical universitaria en un contexto de enseñanza blended-learnmg. *Revista Espanola de Pedagogia*, 75(268), 497–515. <https://doi.org/10.22550/REP75-3-2017-02>
- UNESCO. (n.d.). *Distance Education*. <https://inee.org/eie-glossary/distance-education#:~:text=Distance education is an educational,and time from the learner.>
- Wiley, D. (2006). Innovative Teaching & Learning Strategies (Course/Program Level). *A National Dialogue: The Secretary of Education's Commission On The Future Of Higher Education*. <http://www.ed.gov/news/speeches/2005/09/09192005.html>
- Zainol, Z., Zahari, M. O. I. M., Mazlan, M. H. and Kadir, H. M. S. @ A. (2018). Digital disruption in higher eearning Education: The implementation of blended learning for undergraduates. *Global Business and Management Research: An International Journal*, 10(3), 579–588.



Esta obra está bajo una licencia de Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional.

ChatGPT: La inteligencia artificial como herramienta de apoyo al desarrollo de las competencias STEM en los procesos de aprendizaje de los estudiantes

ChatGPT: Artificial intelligence as a tool supporting the development of STEM skills in student learning processes

Yessica Rodríguez Almazán¹
yessicarodriguez_dpe5@cretam.edu.mx
Ezra Federico Parra-González²
ezra.parra@cimat.mx
Kathia Anahí Zurita-Aguilar^{3*}
kathia.zurita@cusur.udg.mx
Jezreel Mejía Miranda¹
jmejia@cimat.mx
David Bonilla Carranza⁴
jose.bcarranza@academicos.udg.mx

Centro Regional de Formación Docente e Investigación Educativa. Parque Científico y Tecnológico de Tamaulipas, CP 87138, Cd. Victoria, Tamaulipas¹

Centro de Investigación en Matemáticas A. C. Unidad Zacatecas. Parque Quantum, CP 98160. Zacatecas, Zacatecas, México²

Centro Universitario del Sur. Universidad de Guadalajara, CP 49000, Cd. Guzmán Centro, Jalisco, México.³

Centro Universitario del Sur. Universidad de Guadalajara, CP 44430 Guadalajara, Jalisco, Mexico.⁴

RESUMEN

La implementación de chatbots como ChatGPT en la educación STEM presenta beneficios y desafíos. Estas herramientas permiten a los estudiantes desarrollar habilidades tecnológicas y de resolución de problemas, así como mejorar la comunicación y recibir retroalimentación personalizada. En este artículo analizaremos el potencial de la integración de los chatbots en la educación STEM y exploraremos los beneficios que ofrecen, como el fomento del aprendizaje interactivo y personalizado, el desarrollo de habilidades en tecnología y comunicación, y la mejora de la eficiencia en la corrección de tareas. También discutiremos los desafíos que enfrentan los docentes al implementar esta tecnología, como la falta de conocimientos técnicos, la necesidad de formación especializada y las preocupaciones éticas y de privacidad. Proporcionaremos recomendaciones para superar estos desafíos, como ofrecer capacitación docente adecuada, seleccionar herramientas apropiadas y establecer políticas claras de protección de datos. Al finalizar, buscamos destacar la importancia de una implementación adecuada de los chatbots en la educación para maximizar sus beneficios y preparar a los estudiantes para un futuro impulsado por la tecnología.

Palabras clave: Chatbots, ChatGPT, Educación STEM, Inteligencia artificial

ABSTRACT

The implementation of chatbots like ChatGPT in STEM education presents both benefits and challenges. These tools enable students to develop technological and problem-solving skills, as well as enhance communication and receive personalized feedback. This article will examine the potential of integrating chatbots in STEM education and explore the benefits they offer, such as promoting interactive and personalized learning, fostering technology and communication skills, and improving task correction efficiency. We will also discuss the challenges that teachers face when implementing this technology, including lack of technical knowledge, the need for specialized training, and ethical and privacy concerns. We will provide recommendations for overcoming these challenges, such as providing adequate teacher training, selecting appropriate tools, and establishing clear data protection policies. In conclusion, we aim to highlight the importance of proper implementation of chatbots in education to maximize their benefits and prepare students for a technology-driven future.

Keywords: Chatbots, ChatGPT, STEM education, Artificial intelligence.

1. INTRODUCCIÓN

En la actualidad la Inteligencia Artificial (IA) se ha convertido en parte de nuestras vidas, ya que de una u otra forma la utilizamos constantemente a través del uso de aplicaciones. Según Russell y Norvig (2010), la IA se define como "el estudio de cómo hacer que las computadoras realicen tareas que, por el momento, los humanos hacen mejor".

La IA se ha desarrollado a lo largo de varias décadas, sin embargo, se han experimentado avances significativos en las últimas dos décadas gracias al aprendizaje profundo y las redes neuronales artificiales (Goodfellow et al., 2016). Según Domingos (2015), la IA se está convirtiendo en una herramienta cada vez más importante en campos como la medicina, la ingeniería, la física, la biología y las ciencias sociales.

Ejemplo de esto es ChatGPT, que al momento de la redacción del presente trabajo su última versión es la GPT-3, y se trata de un modelo de lenguaje de inteligencia artificial desarrollado por OpenAI, en él se utilizan técnicas de aprendizaje profundo para generar textos coherentes de lenguaje natural en respuesta a una entrada de texto dada y se basa en una arquitectura de red neuronal llamada "Transformer" para generar texto que imita el estilo y el tono del texto de entrada. ChatGPT es un modelo de lenguaje generativo pre-entrenado, que se utiliza para la generación de lenguaje natural. Es capaz de completar textos, traducir idiomas, responder preguntas, entre otras cosas. El modelo consta de 175 mil millones de parámetros, lo que lo convierte en uno de los modelos de inteligencia artificial más grandes y avanzados del mundo. Para entrenar el modelo, se utilizó una gran cantidad de texto de internet, incluyendo artículos de noticias, libros, sitios web, entre otros (Brown et al., 2020).

Los chatbots, como ChatGPT, representan cómo la inteligencia artificial puede ser utilizada para crear sistemas capaces de entender y generar lenguaje natural. Estas aplicaciones tienen un gran potencial en áreas como la comunicación, la educación, la atención médica, etc. Los chatbots conversacionales pueden responder preguntas de los usuarios y llevar a cabo conversaciones naturales, brindando una experiencia interactiva y útil.

2. INTELIGENCIA ARTIFICIAL Y EDUCACIÓN STEM

STEM (por sus siglas en inglés, Ciencia, Tecnología, Ingeniería y Matemáticas), se ha convertido en un término común en la educación y la política de muchos países. Se refiere a un enfoque educativo que fomenta el aprendizaje interdisciplinario en estas áreas.

Una de las definiciones más ampliamente aceptadas de STEM es la propuesta por Velikova et al. (2018), la cual se refiere a STEM como la integración de ciencias naturales, tecnología, ingeniería y matemáticas con el propósito de desarrollar las habilidades de los estudiantes para encontrar soluciones, implementar mejoras e innovaciones a problemas complejos

La NSF definió la educación en STEM como una "integración interdisciplinaria de la ciencia, la tecnología, la ingeniería y las matemáticas en la educación preescolar y primaria, la educación secundaria y la educación superior" (National Science Foundation, 2020).

Una de las ventajas de la metodología STEM es que se enfoca en el aprendizaje práctico y en la aplicación del conocimiento. En lugar de simplemente memorizar hechos y conceptos, los estudiantes aprenden mediante la experimentación, el descubrimiento y la resolución de problemas reales. Esta metodología fomenta el pensamiento crítico, la toma de decisiones y la creatividad, habilidades esenciales para el éxito en cualquier carrera.

Ahora bien ¿cómo se vincula la IA como ChatGPT a la metodología STEM?, como se mencionó en el apartado anterior, la IA se ha ido involucrando en diversos ámbitos de nuestra vida, uno de ellos también es la educación aquí es donde podemos dar respuesta al cuestionamiento planteado, claro que existe una vinculación relevante entre la IA y la metodología STEM en la enseñanza de ambas.

La primera razón es que la inteligencia artificial es una aplicación de la tecnología que utiliza algoritmos y modelos matemáticos para imitar la inteligencia humana en tareas específicas. La creación y mejora de sistemas de inteligencia artificial requiere conocimientos en matemáticas, programación y ciencias de la computación, lo que se relaciona directamente con los principios STEM.

En nuestros tiempos, los chatbots son una de las herramientas más populares de la inteligencia artificial, como sabemos están basados en el procesamiento del lenguaje natural y su dinámica es interactuar con los usuarios mediante una conversación simulada. Dentro de esta categoría, los chatbots de IA como ChatGPT son considerados de los más avanzados, gracias a su capacidad de comprender y responder a una amplia gama de preguntas y respuestas complejas. Por otro lado, la metodología STEM es un enfoque educativo que busca integrar las disciplinas de ciencia, tecnología, ingeniería y matemáticas de una manera más dinámica y aplicada en la enseñanza. El objetivo es fomentar habilidades como la resolución de problemas, la creatividad y el pensamiento crítico en los estudiantes.

En este contexto, los chatbots de IA como ChatGPT pueden ser una herramienta prometedora para fomentar la participación activa de los estudiantes en el aprendizaje de STEM y potenciar su comprensión de los principios científicos y matemáticos (Deng, X., & Yu, Z., 2023).

La capacidad ChatGPT es una oportunidad para interactuar con los estudiantes de manera natural y responder a una amplia variedad de preguntas y respuestas, puede ser especialmente útil para involucrarse en las clases y desarrollar habilidades de STEM en el alumnado. La interacción con chatbots de inteligencia artificial puede resultar en una estrategia eficaz para incrementar el interés y la motivación de los estudiantes hacia las disciplinas STEM. Además, pueden ser especialmente útiles para los estudiantes que tienen dificultades para seguir el ritmo de la enseñanza en las clases de STEM, debido a que puede facilitar la comprensión de los conceptos (Sekeroglu, B., Dimililer, K., & Tuncal, K., 2019).

Como tecnología en constante evolución, es importante seguir explorando nuevas formas de integrar la IA en la educación para mejorar la calidad y eficacia de la enseñanza de STEM. Por otra parte, la implementación de chatbots en la educación de STEM debe ser cuidadosa y bien pensada. La tecnología debe ser utilizada para mejorar la enseñanza y no para reemplazar la interacción y el aprendizaje significativo entre el estudiante y el docente, sin embargo, es crucial realizar una implementación cuidadosa y planificada con el fin de garantizar que los objetivos pedagógicos se cumplan de manera efectiva y se eviten posibles consecuencias negativas, como una dependencia excesiva en las tecnologías (Ding, et al., 2021). En última instancia, la integración de la IA en la educación de STEM debe ser vista como una herramienta complementaria a la educación tradicional, que puede mejorar y enriquecer la experiencia de aprendizaje de los estudiantes.

3. HABILIDADES STEM POSIBLES A DESARROLLAR A TRAVÉS DE CHATGPT

A través de la educación STEM, los estudiantes pueden adquirir habilidades importantes como la resolución de problemas, el pensamiento crítico y la colaboración, entre las que se destacan:

La resolución de problemas se refiere al proceso cognitivo de enfrentar situaciones complejas, identificar los obstáculos y posibles soluciones, y utilizar estrategias de pensamiento y razonamiento para encontrar la mejor respuesta" (Mayer, 1992, p. 25). La educación STEM proporciona a los estudiantes la oportunidad de aplicar el pensamiento crítico y la creatividad para resolver problemas complejos en un entorno práctico. A través de proyectos STEM, los estudiantes pueden aprender a identificar problemas, investigar soluciones y aplicar su conocimiento para encontrar soluciones innovadoras.

La colaboración que implica trabajar en equipo para lograr un objetivo común.

"La colaboración en el enfoque STEM se refiere a la interacción entre estudiantes, maestros y profesionales de diferentes áreas, donde se fomenta el intercambio de conocimientos y habilidades, la resolución conjunta de problemas y la generación de soluciones creativas e innovadoras" (National Research Council, 2011, p. 45).

La educación STEM brinda a los estudiantes la oportunidad de trabajar en proyectos en equipo, lo que les permite desarrollar habilidades importantes como la comunicación, la negociación y la toma de decisiones.

Las habilidades tecnológicas que se pueden desarrollar con la programación, la robótica y la informática. La programación, por ejemplo, les permite a los estudiantes desarrollar habilidades en lógica, pensamiento computacional y resolución de problemas. La robótica les brinda a los estudiantes la oportunidad de aplicar sus habilidades en tecnología para diseñar y construir robots. La informática, por su parte, les proporciona a los estudiantes habilidades en análisis de datos, programación y seguridad informática.

Siendo así los estudiantes que desarrollan estas habilidades a través de la educación STEM están bien preparados para tener éxito en una variedad de carreras y campos. La educación STEM es una inversión importante en el futuro de los estudiantes y puede ayudar a prepararlos para enfrentar los desafíos del mundo moderno. Actualmente, existen varias formas de enseñar y desarrollar habilidades en STEM, y la IA resulta interesante implementarlo en las aulas a través de ChatGPT ya que es una plataforma que se está volviendo cada vez más popular.

Lo anterior conlleva explorar cuáles habilidades STEM y cómo se pueden desarrollar a través de ChatGPT, para ello se citarán estudios relevantes que permiten respaldar esta hipótesis:

En primer lugar, la comprensión de conceptos complejos es una habilidad esencial en STEM y puede ser difícil para muchos estudiantes. La IA a través de ChatGPT puede ayudar a los estudiantes a comprender estos conceptos a través de conversaciones personalizadas y adaptadas a su nivel de comprensión. Los chatbots de IA pueden personalizar el aprendizaje para los estudiantes, ofreciendo una experiencia más individualizada y permitiendo que los estudiantes aprendan a su propio ritmo (Hwang, G.-J., & Chang, C.-Y., 2021). Además, los chatbots pueden proporcionar explicaciones detalladas sobre los conceptos y corregir errores en tiempo real, lo que puede mejorar significativamente la comprensión de los estudiantes.

En segundo lugar, el pensamiento crítico y la resolución de problemas, a través de ChatGPT se puede ayudar a desarrollar estas habilidades por medio de la presentación de problemas y la retroalimentación personalizada. Como afirman Essel et al (2022), "los chatbots de IA pueden presentar problemas complejos a los estudiantes y guiarlos a través del proceso de resolución de problemas, fomentando así el pensamiento crítico y la resolución de problemas". Además, la retroalimentación personalizada proporcionada puede ayudar a los estudiantes a identificar y corregir errores, lo que puede mejorar aún más sus habilidades de pensamiento crítico y resolución de problemas.

En tercer lugar, la colaboración y el trabajo en equipo, ya que muchos proyectos requieren el trabajo en equipo para lograr un resultado exitoso, ChatGPT puede ayudar a los estudiantes a desarrollar estas habilidades a través de la colaboración virtual con compañeros de clase y con los propios chatbots. Como afirman Ding et al. (2021), "los chatbots de IA pueden fomentar la colaboración entre los estudiantes y ayudar a coordinar la comunicación y el trabajo en equipo" (p. 212). Además, también pueden proporcionar retroalimentación personalizada sobre el desempeño individual y del equipo, lo que puede ayudar a los estudiantes a mejorar su colaboración y trabajo en equipo.

El uso de la IA a través de ChatGPT puede ayudar a desarrollar una variedad de habilidades en STEM, incluyendo la comprensión de conceptos complejos, el pensamiento crítico y la resolución de problemas, y la colaboración y el trabajo en equipo. Los estudios citados en este ensayo respaldan la idea de que los chatbots pueden personalizar el aprendizaje para los estudiantes y proporcionar retroalimentación personalizada, lo que puede mejorar significativamente las habilidades en STEM.

4. INTEGRANDO CHATGPT COMO HERRAMIENTA EDUCATIVA

La incorporación de la inteligencia artificial en el aula puede tener un impacto positivo en el aprendizaje de los estudiantes ya que es una herramienta con gran potencial para transformar la educación. Una forma en que se puede involucrar es mediante el uso de asistentes virtuales. Estos asistentes pueden ayudar a los estudiantes a realizar tareas específicas, responder preguntas y proporcionar retroalimentación en tiempo real. Según un estudio realizado por Berggren y Söderström (2021), el uso de asistentes virtuales en el aula puede mejorar la eficacia del aprendizaje, así como la satisfacción y la motivación de los estudiantes.

Otra forma es el uso de sistemas de tutoría inteligente para proporcionar retroalimentación y sugerencias personalizadas a los estudiantes durante su proceso de aprendizaje. Según un estudio realizado por Smith (2019) "La inteligencia artificial ha demostrado su capacidad para proporcionar retroalimentación inmediata y personalizada a los estudiantes, lo que facilita un aprendizaje más individualizado y efectivo" Así mismo Johnson et al (2020) mencionan que "Los tutores virtuales basados en inteligencia artificial pueden adaptarse a las necesidades específicas de cada estudiante, brindando un apoyo constante y personalizado" por lo tanto los sistemas de IA pueden funcionar como una tutoría inteligente y mejorar significativamente el aprendizaje de los estudiantes en comparación con los métodos de enseñanza tradicionales.

También se puede involucrar el ChatGPT mediante la creación de tutoriales y materiales educativos interactivos para generar contenido educativo de alta calidad, que puede ser presentado en formato de video, presentación o texto. Según un estudio realizado por Arana (2021) presenta las áreas que aún no han sido explotadas de la IA la cual una de ellas es la creación de materiales educativos, el propósito consiste en apoyar a los docentes en la creación y mejora de recursos educativos que se ajusten a las características y objetivos de las distintas disciplinas y conocimientos que se desean transmitir, aprovechando las capacidades y características de la inteligencia artificial, en particular, las redes neuronales generativas, por lo tanto, con la ayuda del ChatGPT se puede mejorar significativamente el aprendizaje de los estudiantes.

Siendo más específicos la implementación de GPT en el aula es muy variada depende de cada campo formativo o asignatura que se quiera abordar, por ejemplo: Puede ser utilizado para mejorar el aprendizaje de la lengua española en la clase de lenguaje y comunicación. El chatbot puede ser programado para corregir la ortografía y la gramática de los estudiantes y para proporcionar retroalimentación inmediata sobre su escritura.

En la clase de Ciencias Naturales, se puede usar para responder preguntas de los estudiantes sobre diferentes temas, como la biología, la química y la física. Por ejemplo, puede ser programado para responder preguntas sobre la estructura celular, la tabla periódica y las leyes del movimiento. Además, puede ser programado para proporcionar explicaciones y ejemplos para ayudar a los estudiantes a entender los conceptos científicos. En física, ChatGPT podría ser utilizado para ayudar a los estudiantes a entender mejor los conceptos abstractos. Por ejemplo, podría generar ejemplos de fenómenos físicos en tiempo real y explicar cómo se aplican las leyes físicas a esos fenómenos. También podría ser utilizado para explicar conceptos más abstractos, como la teoría de la relatividad de Einstein, y para ayudar a los estudiantes a entender cómo esos conceptos se aplican en el mundo real.

En química, ChatGPT podría ser utilizado para ayudar a los estudiantes a entender mejor las reacciones químicas. Por ejemplo, podría generar visualizaciones en tiempo real de las reacciones químicas y explicar cómo se producen esas reacciones. También podría ser utilizado para generar preguntas personalizadas para los estudiantes, de acuerdo con sus habilidades y conocimientos previos, y para proporcionar explicaciones adicionales y ejemplos para ayudar a los estudiantes a comprender mejor los conceptos químicos.

En la clase de Matemáticas para ayudar a los estudiantes a resolver problemas. Puede ser programado para proporcionar explicaciones detalladas sobre cómo resolver problemas matemáticos, así como para proporcionar ejemplos y consejos para ayudar a los estudiantes a comprender mejor los conceptos matemáticos, ChatGPT podría trabajar con los estudiantes para ayudarles a entender cómo se llega a la solución correcta. Por ejemplo, en lugar de simplemente proporcionar la respuesta correcta a un problema, Además, también podría ser utilizado para generar problemas personalizados para los estudiantes, de acuerdo con sus habilidades y conocimientos previos, y para proporcionar explicaciones detalladas y ejemplos adicionales para ayudar a los estudiantes a comprender mejor los conceptos matemáticos.

En la clase de Tecnología, puede ser utilizado para proporcionar información sobre diferentes herramientas y tecnologías. Por ejemplo, puede ser programado para proporcionar información sobre el software de diseño gráfico, la programación y el diseño de páginas web. Además, puede ser programado para responder preguntas de los estudiantes sobre diferentes herramientas y tecnologías y ayudarles a comprender cómo utilizarlas.

En la clase de inglés puede utilizarse de diversas formas desde asistente virtual de conversación en inglés hasta corrección gramatical y ortográfica, generación de respuestas y ejemplos, y práctica de lectura en inglés, puede ayudar a los alumnos a mejorar sus habilidades en el idioma inglés de manera interactiva y personalizada.

El uso del chatbots en diferentes asignaturas de secundaria puede ayudar a los estudiantes a desarrollar habilidades STEM, incluyendo la resolución de problemas, la colaboración y la comunicación. El chatbot puede ser utilizado como una herramienta de colaboración en la que los estudiantes trabajan juntos para programar el chatbot y mejorar su conocimiento sobre diferentes temas. Además, puede ser utilizado para fomentar la resolución de problemas y la creatividad, ya que los estudiantes pueden programar el chatbot para responder preguntas de manera única y creativa. La utilidad y aplicación de este resulta variada la cual dependería de la creatividad del docente.

5. RETOS PARA LOS DOCENTES AL IMPLEMENTAR HERRAMIENTAS DE IA

La implementación de la IA en el aula es un tema relevante en la actualidad, y se espera que en un futuro no muy lejano esta tecnología se convierta en una herramienta imprescindible en la educación. Si bien la IA ofrece una serie de beneficios potenciales para la enseñanza y el aprendizaje, también presenta una serie de desafíos y preocupaciones para los docentes. En este sentido, el papel del docente se convierte en un elemento crucial para garantizar una implementación adecuada de en el aula. No obstante, la introducción de esta tecnología en la educación presenta una serie de retos que deben ser abordados para que su impacto sea positivo.

Uno de los principales retos que enfrentan los docentes al implementar la IA en el aula es la falta de conocimiento y experiencia en la tecnología. Como señala Ocaña et al. (2019), muchos docentes pueden no estar familiarizados con las tecnologías de aprendizaje automático o no tener experiencia en el diseño de cursos que involucren tecnologías de IA. Esto puede dificultar la integración efectiva de la IA en el aula, ya que los docentes pueden sentirse abrumados por la tecnología y no saber cómo usarla para mejorar el aprendizaje de sus estudiantes.

Así mismo se enfrentan al cambio tecnológico. Como señala Rodrigo Martín-Romo, profesor de la Universidad Politécnica de Madrid, "la introducción de tecnologías en la enseñanza puede resultar en un cambio significativo en el enfoque pedagógico del docente, lo que implica un proceso de adaptación al nuevo entorno y a las nuevas formas de aprendizaje" (Martín-Romo, 2020).

También los docentes se enfrentan a la falta de formación y capacitación adecuada. Según Picciano (2019), los docentes deben estar preparados para usar la IA de manera efectiva y ética, y esto implica una formación especializada en este campo. Asimismo, deben ser capaces de integrar la IA en su práctica pedagógica de manera efectiva, lo que requiere una comprensión profunda de las tecnologías de la información y la comunicación (TIC) y su impacto en la educación.

Otro desafío es la selección y uso adecuado de las herramientas de IA en el aula. La elección de las herramientas de IA debe basarse en la comprensión de las necesidades de los estudiantes y en la capacidad de la tecnología para mejorar la calidad de la enseñanza. Según la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO, por sus siglas en inglés] (2020), el uso de la IA en la educación debe ser diseñado para ayudar a los estudiantes a desarrollar habilidades cognitivas y sociales, y no para reemplazar al docente.

Además, la implementación de la IA en el aula plantea desafíos éticos y de privacidad que deben ser abordados adecuadamente. Según Perkins (2019), los docentes deben garantizar que la información recopilada por las herramientas de IA se use de manera responsable y no se utilice para fines inapropiados, como el seguimiento y la evaluación de los estudiantes. Asimismo, deben garantizar la privacidad de los estudiantes y protegerlos de posibles violaciones de la seguridad de datos. Como señala ProFuturo y OEI (2023) los docentes tienen la responsabilidad ética de proteger la privacidad de los datos de sus estudiantes. La recopilación y el análisis de datos de los estudiantes por parte de la IA puede generar preocupaciones en relación con la privacidad y la seguridad de los datos, lo que puede hacer que los docentes se sientan incómodos con la tecnología.

Otro reto importante que enfrentan los docentes al implementar la IA en el aula es la necesidad de adaptar la tecnología a las necesidades y habilidades de sus estudiantes. Como señala Hargreaves (2018), "los docentes deben ser capaces de adaptar la IA para satisfacer las necesidades de sus estudiantes individuales" (p. 15). Esto puede requerir una comprensión profunda de las necesidades y habilidades de cada estudiante, lo que puede ser difícil de lograr en un entorno de clase lleno de estudiantes.

Por último, la IA puede impactar en la relación entre el docente y el estudiante. Como señala David Perkins, profesor emérito de la Universidad de Harvard, "la IA puede ser útil en ciertos aspectos del aprendizaje, pero no puede reemplazar la relación personal y significativa entre el docente y el estudiante" (Perkins, 2019). Los docentes deben asegurarse de que la IA no reemplace esta relación y sigan involucrados en el proceso de aprendizaje de sus estudiantes.

Como vemos la integración de la IA en el entorno educativo presenta diversos retos para los docentes, quienes deben estar preparados para adaptarse a esta nueva forma de enseñanza para ello es importante que los docentes se eduquen y se preparen adecuadamente para enfrentar estos desafíos y aprovechar al máximo los beneficios potenciales de la IA en la educación.

6. RECOMENDACIONES PARA SUSPERAR LOS RETOS

Para superar los retos que enfrentan los docentes al implementar la IA en el aula, es necesario que se les brinde una formación especializada y capacitación adecuada. Esto puede lograrse a través de programas de capacitación en línea y presenciales que permitan a los docentes adquirir habilidades técnicas y pedagógicas para integrar la IA en su práctica educativa de manera efectiva.

Además, es importante que los docentes seleccionen las herramientas de IA adecuadas para su práctica pedagógica. Para ello, es necesario que los docentes estén familiarizados con las diferentes opciones disponibles y evalúen las herramientas en función de su capacidad para mejorar la enseñanza y el aprendizaje.

Por último, es necesario que los docentes tomen medidas para proteger la privacidad de los estudiantes y garantizar el uso ético de la IA en el aula. Esto implica el establecimiento de políticas claras y prácticas adecuadas para el manejo y almacenamiento de datos de los estudiantes, así como la promoción de la responsabilidad y la ética en el uso de la IA en la educación.

La IA es una tecnología que tiene el potencial de transformar la educación, pero su implementación en el aula presenta una serie de desafíos que deben ser abordados para que su impacto sea positivo. En este sentido, los docentes tienen un papel crucial en la introducción de la IA en la educación, y deben estar preparados para integrar de manera efectiva y ética.

7. CONSIDERACIONES FINALES

La educación STEM es esencial en el mundo actual, donde la tecnología está transformando rápidamente todos los ámbitos de la sociedad. En este contexto, la implementación de la tecnología de procesamiento de lenguaje natural, como los chatbots, puede ser beneficiosa para los estudiantes de secundaria al fomentar el aprendizaje y el desarrollo de habilidades STEM. La tecnología está transformando la educación en todo el mundo y la inteligencia artificial es una de las tecnologías que están liderando el cambio.

Una de las ventajas de utilizar chatbots en la educación es que pueden ayudar a los estudiantes a desarrollar habilidades en el campo de la tecnología y las ciencias de la computación. Los estudiantes pueden interactuar con los chatbots en tiempo real y aprender de una manera más interactiva y personalizada. Los chatbots pueden proporcionar una respuesta inmediata a las preguntas de los estudiantes, lo que les permite aprender de manera autónoma y a su propio ritmo. Además, los chatbots pueden ser programados para enseñar diferentes temas y asignaturas, lo que los convierte en una herramienta muy versátil en el aula.

Otra ventaja importante de la implementación de chatbots en la educación es que pueden ayudar a los estudiantes a desarrollar habilidades en el campo de la comunicación y la interacción social. Al interactuar con chatbots, los estudiantes pueden aprender a comunicarse de manera más efectiva y clara. También pueden mejorar su capacidad para escuchar y responder a los demás, lo que es esencial para el trabajo en equipo y la colaboración en proyectos de grupo. Los chatbots GPT pueden proporcionar una forma segura y cómoda de practicar habilidades de comunicación y socialización sin el temor de ser juzgados.

Otra ventaja importante de los chatbots es que pueden mejorar la eficiencia en el aprendizaje y reducir el tiempo dedicado a la corrección de tareas. Los chatbots pueden ser programados para corregir y evaluar tareas, proporcionando una respuesta inmediata y precisa a los estudiantes. Esto permite a los docentes dedicar más tiempo a la instrucción y a la interacción con los estudiantes, lo que mejora la calidad del aprendizaje y la experiencia de los estudiantes.

Uno de los principales beneficios en la educación es la capacidad de proporcionar retroalimentación inmediata y personalizada a los estudiantes. Los chatbots pueden analizar el rendimiento del estudiante y proporcionar sugerencias específicas para mejorar su aprendizaje. Además, pueden ser programados para adaptarse a las necesidades individuales de cada estudiante, proporcionando un enfoque más personalizado que puede ser especialmente beneficioso para los estudiantes con diferentes niveles de habilidad y de aprendizaje.

Otra ventaja es que pueden ayudar a los estudiantes a desarrollar habilidades de pensamiento crítico y resolución de problemas. Los chatbots pueden ser programados para proporcionar a los estudiantes problemas desafiantes y preguntas, y ayudarlos a encontrar soluciones creativas y efectivas. Al trabajar con un chatbot, los estudiantes también pueden aprender a comunicarse de manera clara y efectiva, lo que puede ser útil en muchos campos de STEM.

También pueden ser utilizados para ayudar a los estudiantes a explorar diferentes áreas de STEM y proporcionar una experiencia de aprendizaje más interactiva y atractiva. Los chatbots pueden ser programados para proporcionar información sobre temas de ciencia y tecnología, y pueden ser utilizados para crear simulaciones y juegos educativos. Esto puede ayudar a los estudiantes a desarrollar su curiosidad y motivación para aprender sobre temas STEM.

Para los docentes, pueden ser una herramienta útil para reducir la carga de trabajo y ahorrar tiempo. Los chatbots pueden ayudar a corregir tareas y proporcionar retroalimentación a los estudiantes, lo que permite a los docentes dedicar más tiempo a la instrucción y la interacción con los estudiantes. También pueden ser programados para realizar tareas repetitivas, lo que reduce la carga de trabajo de los docentes.

A manera de conclusión, ChatGPT puede ser una herramienta valiosa en el aula para mejorar el aprendizaje y la experiencia de los estudiantes. Es importante que los docentes conozcan y utilicen esta tecnología en su enseñanza para mejorar la calidad de la educación y preparar a los estudiantes para un futuro impulsado por la tecnología. Además, pueden ahorrar tiempo y reducir la carga de trabajo de los docentes, lo que les permite enfocarse en la instrucción y la interacción con los estudiantes.

8. REFERENCIAS

- Arana, C. (2021). Inteligencia Artificial Aplicada a la Educación: Logros, Tendencias y Perspectivas. *INNOVA UNTREF. Revista Argentina de Ciencia y Tecnología*.
- Berggren, A. & Söderström, T. (2021). Virtual assistants in higher education - from research to practice. *Education and Information Technologies*, 26(3), 2865-2883.
- Blikstein, P. (2018). Artificial intelligence in education promises and implications for teaching and learning. *Journal of Educational Technology Development and Exchange (JETDE)*, 11(1), 1-14.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- Deng, H., Chen, X., Jozefowicz, R., Lau, J. H., Shazeer, N., & Smith, N. A. (2021). GPT-2 1.5B Release. OpenAI Blog.
- Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability*, 15(4), 2940.
- Ding, X., et al. (2021). An Intelligent Tutoring System with Chatbot: Effects on Student Learning, Engagement, and Satisfaction. *IEEE Transactions on Learning Technologies*, 14(2), 212-223.
- Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books.
- Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. *International Journal of Educational Technology in Higher Education*, 19(1), 1-19.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Hargreaves, A. (2018). The fourth way of digital technology and the (in) human future of teaching. *European Educational Research Journal*, 17(1), 13-28.
- Hwang, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*. Advance online publication.
- Jang, S., Kim, S., & Kang, S. (2021). Development of Interactive Educational Contents Based on Deep Learning Generative Models. *IEEE Access*, 9, 133675-133686.
- Johnson, R., Williams, K., & Lee, S. (2020). Artificial Intelligence as a Virtual Tutor: Current Trends and Future Directions. *Educational Technology Review*, 28(4), 567-584.
- Martín-Romo, R. (2020). The use of Artificial Intelligence in Education: Challenges and Opportunities. *International Journal of Educational Technology in Higher Education*, 17(1), 1-9.
- Mayer, R. E. (1992). *Thinking, problem solving, cognition*. WH Freeman/Times Books/Henry Holt & Co.
- National Research Council. (2011). *Successful K-12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics*. National Academies Press.
- National Science Foundation. (2020). *STEM Education of future*. Recuperado de <https://www.nsf.gov/edu/Materials/STEM%20Education%20for%20the%20Future%20-%202020%20Visioning%20Report.pdf>
- Ocaña-Fernández, Y., Valenzuela-Fernández, L. A., & Garro-Aburto, L. L. (2019). Inteligencia artificial y

- sus implicaciones en la educación superior. *Propósitos y representaciones*, 7(2), 536-568.
- Perkins, D. (2019). Artificial Intelligence and the Future of Education. *Educational Researcher*, 48(4), 216-222.
- Perkins, D. (2019). The ethics of AI in education: Mapping the debate. *Education Sciences*, 9(1), 35.
- Picciano, A. G. (2019). Artificial intelligence and machine learning in education. In S. B. Merriam & R. S. Courtenay (Eds.), *The Jossey-Bass handbook of innovative higher education* (pp. 307-323). John Wiley & Sons.
- ProFuturo y OEI. (2023). El futuro de la Inteligencia Artificial en educación en América Latina.
- Russell, S. J., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (Third Edition). Pearson.
- Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019). La Inteligencia Artificial en Educación: aplicación en la evaluación del desempeño del alumno. *Dilemas contemporáneos: Educación, Política y Valores*.
- Smith, A. (2019). Artificial Intelligence in Education: Promises and Implications. *Journal of Educational Technology*, 43(2), 356-371.
- United Nations Educational, Scientific and Cultural Organization. (2017). *STEM Education for Sustainable Development: A Learning Module*. Recuperado de <https://unesdoc.unesco.org/ark:/48223/pf0000247453>
- United Nations Educational, Scientific and Cultural Organization. (2020). *Artificial intelligence and education: Challenges and opportunities for sustainable development*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000373821>
- Velikova, E., Mierlus-Mazilu, I., Vasileva-Ivanova, R., & Georgieva, D. (2018). About the stem Education. *Proceedings of University of Ruse*, 57, 10-14.



Esta obra está bajo una licencia de Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Recibido 23 May. 2023

ReCIBE, Año 12 No. 1, May. 2023

Aceptado 05 Jul. 2023

Revisión de Problemas en la Detección de Objetos en Imágenes y Videos Digitales

A Survey on Problems of Detection of Objects in Digital Images and Videos

Miguel Ángel Gutiérrez Velázquez¹
m19061419@chihuahua.tecnm.mx

Mario Ignacio Chacon-Murguia¹
mario.cm@chihuahua.tecnm.mx

Alma Delia Corral-Saenz¹
alma.cs@chihuahua.tecnm.mx

¹ Tecnológico Nacional de México / I.T. Chihuahua, Chihuahua, Chihuahua, México.

Resumen

En las últimas décadas, la detección de objetos ha sido una tarea muy importante en el área de visión por computadora, ya que la detección de objetos localiza y clasifica uno o más objetos en una imagen o videos. En este artículo, se presenta una revisión de artículos y se describen técnicas clásicas y de aprendizaje profundas utilizadas para la detección de objetos. Además, se realiza una revisión de trabajos recientes sobre la detección de objetos, enfocándose en cómo se solucionan algunos de sus problemas más relevantes. Los problemas que se abarcan son: oclusión, confusión, información contextual, cambios en la iluminación, objetos pequeños y cambios de escala, variación entre la misma clase y diferentes clases, y deformación y cambios de pose. Se espera que este artículo sirva para que los interesados en el área conozcan ideas y enfoques para resolver problemas existentes en la detección de objetos incluyendo los últimos avances.

Abstract

In recent decades, object detection has been a very important task in the area of computer vision, since object detection locates and classifies one or more objects in an image or video. In this article, a review of articles is presented, and classical and deep learning techniques used for object detection are described. In addition, a review of recent works on object detection is carried out, focusing on how some of its most relevant problems are solved. Issues covered are occlusion, confusion, contextual information, lighting changes, small objects and scale changes, variation between the same class and different classes, and deformation and pose changes. It is hoped that this article will help those interested in the area to learn about ideas and approaches to solving existing problems in object detection, including the latest advances.

Palabras clave: Detección de objetos, Problemas en la detección de objetos, Aprendizaje profundo.

Keywords: Object detection, Problems in object detection, Deep Learning.

1. Introducción

La detección de objetos es parte fundamental del proceso de la visión por computador, ya que brinda información de interés presente en una imagen. La detección de objetos es un área en constante evolución, lo que hace fundamental el mantenerse actualizado en los nuevos métodos e ideas relacionados a ella.

El campo de aplicación de la detección de objetos es amplio: detección de animales (Wang L., *et al.*, 2021; Li N., *et al.*, 2020; Yudin D., *et al.*, 2019; Kellenberger B., *et al.*, 2019; Singh A., *et al.*, 2020), detección de peatones (Han. B., *et al.*, 2020; Cygert S. y Czyzewski A., 2020), de billetes (Rodríguez A., *et al.*, 2020), diagnóstico médico (Gurbina M., *et al.*, 2019; Pathare S., *et al.*, 2020; Meda K., *et al.*, 2021), detección de desastres (Muhammad K., *et al.*, 2018; Radhika S., *et al.*, 2018), astronomía (Wu T., 2020), agricultura (Song C., *et al.*, 2020; Kaur M. y Min C., 2018), detección de objetos bajo el agua (Chen Z., *et al.*, 2020), detección de medios de transporte (Mo Y., *et al.*, 2019; Yilmaz B. y Karsligil M., 2020; Huang G., *et al.*, 2019), detección del horizonte (Zardoua Y., *et al.*, 2021), deportes (Guo. T., *et al.*, 2020; Bastanfard A., *et al.*, 2019), etcétera. Sin embargo, a pesar de ser un área ampliamente estudiada durante varias décadas, la detección de objetos es un concepto que carece de un acuerdo universal (Liu L., *et al.*, 2019) ya que hay autores que limitan el alcance a la localización de objetos (Pandiya M., *et al.*, 2020), otros primero realizan una detección y después la localización de los objetos (Pathak A., *et al.*, 2018; Adreopoulos A. y Tsotsos J., 2013) y otros más, consideran la clasificación de los mismos como parte del proceso de detección (Vashisht M. y Kumar B., 2020; Zhang H. y Hong X., 2019). Asimismo, se incorporan variantes a estas etapas como realizar la localización mediante rectángulos delimitadores (Biswas S., *et al.*, 2021; Xiao Y., *et al.*, 2020) o bien, determinar primero si existen objetos de interés y de ser así, regresar la localización y la etiqueta de la categoría (Liu L., *et al.*, 2019).

Además de los enfoques de detección de objetos, existen trabajos publicados que describen posibles soluciones a problemas o aplicaciones específicas en la detección de objetos, como los mostrados en la Tabla . Los trabajos también presentan arquitecturas, métricas y conjuntos utilizados. La Figura 1 muestra de manera sistemática el contenido y organización de este artículo.

2. Análisis de estudios existentes

Debido a la gran cantidad de trabajos que hay en la literatura sobre detección de objetos, es de utilidad hacer una recopilación de los artículos más significativos en los avances del tema, sus métodos y las aplicaciones específicas.

A partir del análisis de los artículos se observaron las siguientes tendencias: análisis de aplicaciones específicas de detección de objetos, detección de objetos 3D, detección de objetos sobresalientes en una imagen, detección de objetos en imágenes RGB-D, detección de objetos pequeños, detección de objetos camuflados, detección de objetos en movimiento, detección de objetos con Few-Shot, detección de objetos con YOLO y análisis general del área de detección de objetos (Shetty A., *et al.*, 2021; Deng J., *et al.*, 2020; Kaushal M., *et al.*, 2018; Aziz L., *et al.*, 2020; Jiao L., *et al.*, 2019; Arulprakash E. y Aruldoss M., 2021; Abbas S., *et al.*, 2022; Zou Z., *et al.*, 2023); siendo estos artículos de análisis generales los recomendados para aquellos lectores que deseen introducirse al área de detección de objetos.

En la Tabla se presenta una descripción más detallada de los artículos mencionados en esta sección. Con base en el análisis previo, este artículo analiza y describe las ideas y enfoques implementados para resolver problemas presentados frecuentemente en la detección de objetos. Además, hace un hincapié en posibles aportes para mejorar y hacer más robusta la detección de objetos ante los problemas mencionados en este artículo, así como recomendaciones para una futura investigación.

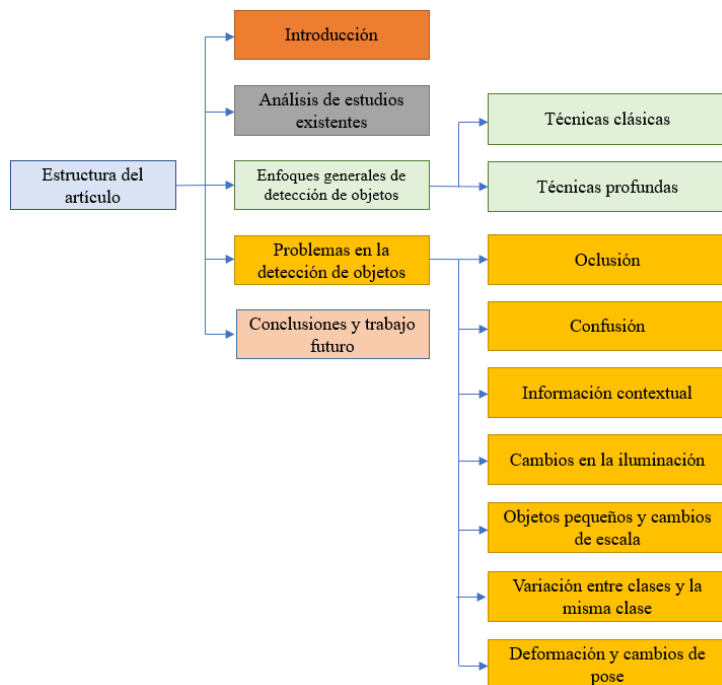


Figura 1. Estructura del artículo.

Artículo	Tema	Resumen
(Mohammed S., et al., 2021)	Detección de barcos	Analizan dos algoritmos: Multi-Input Convolutional Long Short Term Memory y Rotational Libra Region Based CNN para la detección de barcos.
(Moniruzzaman M., et al., 2017)	Detección de objetos debajo del agua	Mencionan detectores profundos en la detección de objetos marinos, así como enfoques a objetos específicos, como pescados, plancton o corales.
(Xu S., et al., 2023)	Detección de objetos debajo del agua	Presenta un análisis de los problemas en la detección de objetos bajo el agua, así como futuras tendencias y aplicaciones. Además, se analiza la relación entre el mejoramiento de imágenes y la detección de objetos.
(Ajmera F., et al., 2021)	Detección de objetos en imágenes aéreas	Resumen de varios métodos tanto de aprendizaje profundo como con enfoques clásicos para la detección de objetos en imágenes aéreas. También se mencionan las ventajas, desventajas, precisión y conjuntos de datos.
(Mittal P., et al., 2020)	Detección de objetos en imágenes aéreas	Se hace una revisión de algoritmos de aprendizaje profundo recientes en la detección de objetos en imágenes tomadas con UAV (Unmanned Aerial Vehicle) a baja altitud.
(Said N., et al., 2019)	Detección de desastres	Revisión de detección de desastres en redes sociales y en imágenes satelitales. Además, se discuten trabajos futuros y desafíos existentes.
(Geetha S., et al., 2021)	Detección de fuego	Mencionan métodos fundamentales de procesamiento de imágenes y CNN para la detección de humo y fuego, también conjuntos de datos existentes y desafíos y sugerencias de mejoras.
(Liang D., et al., 2020)	Detección de líneas	Organizan métodos dependiendo en los objetos a detectar y se presenta un nuevo conjunto de datos.
(Kaur P., et al., 2020)	Detección de tumores	Revisan varias técnicas de detección de tumores en imágenes MR junto con sus dificultades y fortalezas. Se cubren las estrategias más relevantes, restricciones, métodos, sus reglas de aplicación, y las direcciones futuras.

Tabla 1. Comparación de artículos de estudios sobre la detección de objetos.

Artículo	Tema	Resumen
(Naji S., et al., 2019)	Detección de piel	Se revisan métodos para la detección de la piel en imágenes a color. Además, se mencionan espacios de color, costos, riesgos, bases de datos, pruebas y una evaluación comparativa.
(Feng Y., et al., 2021)	Detección de rostros	Se presentan métodos, análisis en términos de precisión y eficiencia de detectores profundos de rostros. Se comparan conjuntos de datos y métricas de evaluación.
(Minaee S., et al., 2021)	Detección de rostros	Se hace revisión de métodos profundos para la detección de rostros y se proponen categorías de agrupamiento, se presentan arquitecturas y sus desempeños en evaluaciones comparativas populares, se presentan conjuntos de datos y se mencionan desafíos actuales y futuras direcciones de investigación.
(Ning C., et al., 2021)	Detección de peatones	Análisis del progreso en la detección de peatones con oclusión.
(Espinosa J., et al., 2020)	Detección de motocicletas	El artículo se enfoca en los algoritmos para la detección y seguimiento de motocicletas, se describe el desempeño en conjuntos de datos, además se presentan desafíos actuales, así como trabajo futuro.
(Jarunakarint V., et al., 2020)	Detección de motocicletas	Realizan experimentos en 20 modelos distintos de aprendizaje de máquina para detectar motocicletas.
(Nyein M. y Tint T., 2021)	Detección de vehículos	Se hace un análisis del desempeño de diversos métodos para la detección de vehículos mediante resultados experimentales utilizando conjuntos de datos habituales.
(Li Z., et al., 2021)	Detección de objetos 3D	Revisión del desarrollado en la detección de objetos 3D aplicados a la conducción inteligente, así como sus defectos y futuras investigaciones.
(Lang W., et al., 2021)	Detección de objetos 3D	Se enfocan en métodos populares de aprendizaje profundo para la detección de objetos 3D. Se dividen los enfoques en 4 categorías de acuerdo con los datos de entrada. Se hace un análisis de las innovaciones y se hace una comparación de sus desempeños. Así como los desafíos actuales y futuras direcciones.
(Arnold E., et al., 2019)	Detección de objetos 3D	Se revisan métodos utilizados en la conducción autónoma para la detección de objetos 3D. Se categorizan trabajos recientes basados en el tipo de sensor. Se indican problemas actuales y trabajos futuros.
(Kumar N., 2018)	Detección de objetos sobresalientes	Análisis de varios métodos de umbralización utilizados en el estado del arte en SOD. Además, se mencionan umbrales no explorados, se realizan, también, experimentos para mostrar el desempeño de umbralizaciones populares.
(Wang Q., et al., 2020)	Detección de objetos sobresalientes en videos	Hacen una clasificación de métodos del estado del arte, se mencionan los conjuntos de datos y las métricas de evaluación más utilizadas, se realizan experimentos para comparar el desempeño de los métodos y, finalmente, se mencionan trabajos futuros.
(Ullah I., et al., 2020)	Detección de objetos sobresalientes	Se hace una revisión de técnicas heurísticas y de aprendizaje profundo, además de campos relacionados a SOD (Eye-fixation-prediction, detección en imágenes RGB-D, detección de objetos coprominentes, en video). Además, se mencionan conjuntos de datos y problemas actuales y trabajos futuros.
(Wang W., et al., 2021)	Detección de objetos sobresalientes	Revisan algoritmos de SOD: su arquitectura, nivel de supervisión, paradigma de aprendizaje, detección nivel objeto/instancia. También métricas y conjuntos de datos, comparación de modelos SOD representativos. Se expone un conjunto de datos desafiante. Se analiza el desempeño de modelos SOD ante perturbaciones en la entrada y ataques adversarios. Se mencionan problemas actuales y direcciones para investigaciones futuras.
(Borji A., et al., 2019)	Detección de objetos sobresalientes	Muestran el progreso en SOD, sus inicios, conceptos fundamentales, las técnicas principales y conjuntos de datos y métricas de evaluación.

Tabla 1. Comparación de artículos de estudios sobre la detección de objetos [continuación].

Artículo	Tema	Resumen
(Borji A., et al., 2019)	Detección de objetos sobresalientes	Muestran el progreso en SOD, sus inicios, conceptos fundamentales, las técnicas principales y conjuntos de datos y métricas de evaluación.
(Han J., et al., 2018)	Detección de objetos sobresalientes	Se Presentan definiciones, tareas, técnicas recientes, investigaciones esenciales, conjuntos de datos y métricas de evaluación y comparación y análisis de resultados experimentales. Y se hacen hincapié en la relación entre detección de objetos genéricos, sobresalientes y específicos a una clase.
(Zhou T., et al., 2021)	Detección de objetos sobresalientes	Se hace una revisión de modelos SOD enfocados a imágenes RGB-D, así como conjuntos de datos para ese dominio. Se hace una evaluación de diversos métodos. Finalmente, se mencionan desafíos existentes para investigaciones futuras.
(Liu Y., Sun P., et al., 2021)	Detección de objetos pequeños	Resumen de métodos de aprendizaje profundo para la detección de objetos pequeños. Se resumen desafíos y soluciones. Se incorporan técnicas de cuatro áreas de investigación: detección de objetos genéricos, detección de rostros, detección de objetos en imágenes aéreas y segmentación. Se hace comparación de diversos métodos de aprendizaje profundo utilizando tres bases de datos.
(Tong K., et al., 2020)	Detección de objetos pequeños	Revisión de métodos basados en aprendizaje profundo desde cinco aspectos: aprendizaje de características multiescala, aumento de datos, estrategias de entrenamiento, detección basada en el contexto y detección basada en redes tipo GAN. Se hace un análisis de algunos algoritmos típicos en los conjuntos de datos MS-COCO y PASCAL-VOC. Se concluye con posibles investigaciones futuras.
(Cheng G., et al., 2022)	Detección de objetos pequeños	Se efectúa un análisis de la literatura sobre la detección de objetos pequeños. Además, se presentan dos conjuntos de datos para la detección de objetos pequeños en escenarios aéreos y de conducción.
(Mondal A., 2021)	Detección de objetos camuflajeados	Análisis de técnicas utilizando procesamiento de imágenes, aprendizaje de máquina, gráficos de computadora y aprendizaje profundo. Se presentan conjuntos de datos y se mencionan futuras investigaciones posibles.
(Chapel M. y Bouwmans T., 2018)	Detección de objetos en movimiento	Revisión de métodos basados en la substracción del fondo, clasificación de trayectoria, matriz de descomposición y seguimiento de objetos de objetos en movimiento en secuencias de videos adquiridas por una cámara en movimiento.
(Chapel M. y Bouwmans T., 2020)	Detección de objetos en movimiento	Se Analizan artículos sobre la detección de objetos en movimiento en videos tomados por una cámara en movimiento. Se proponen categorías para identificar los métodos existentes en el estado del arte de acuerdo con la representación de la escena. También se abordan métodos y enfoques para cámaras sin movimiento, conjuntos de datos y métricas de evaluación.
(Huang G., et al., 2023)	Detección de objetos con Few-Shot	Análisis y caracterización de detección de objetos con Few-Shot y aprendizaje autosupervisado.
(Köhler M., et al., 2023)	Detección de objetos con Few-Shot	Revisión del estado del arte sobre la detección de objetos con Few-Shot; así como conjuntos de datos utilizados, evaluaciones y comparativa de resultados. Además, se mencionan los problemas frecuentes y las posibles tendencias futuras.
(Diwan T., et al., 2023)	Detección de objetos con YOLO	Se presenta un análisis del detector YOLO; su evolución en las arquitecturas, comparaciones, aplicaciones y posibles investigaciones futuras.
(Terven J. y Cordova D., 2023)	Detección de objetos con YOLO	Análisis de las arquitecturas YOLO. Desde la primera versión de YOLO hasta YOLOv8.

Tabla 1. Comparación de artículos de estudios sobre la detección de objetos [continuación].

Artículo	Tema	Resumen
(Shetty A., et al., 2021)	Modelos de detección de objetos	Se realiza un análisis comparativo entre técnicas basadas en aprendizaje profundo (de una y dos etapas) para la detección de objetos.
(Deng J., et al., 2020)	Modelos de detección de objetos	Revisan modelos profundos (de una y dos etapas), se reportan conjunto de datos y desafíos existentes.
(Kaushal M., et al., 2018)	Modelos de detección de objetos	Se hace una recopilación de trabajos con aprendizaje profundo, lógica difusa, algoritmos evolutivos e híbridos para la detección y seguimiento de objetos. Se remarcan los conjuntos de datos. Se presentan desafíos existentes y análisis que pueden guiar investigaciones futuras.
(Aziz L., et al., 2020)	Modelos de detección de objetos	Se revisan modelos de una y dos etapas. Se presentan cinco conjuntos de datos con sus métricas de evaluación. El artículo se centra en cinco aplicaciones: seguridad, milicia, transporte, medicina y vida diaria.
(Jiao L., et al., 2019)	Modelos de detección de objetos	Analizan detectores de una y dos etapas. Se presentan conjuntos de datos, se realiza, también, un análisis de métodos generales para mejorar la detección de objetos. Se mencionan diversas aplicaciones y, finalmente, se presentan tendencias y desafíos interesantes.
(Arulprakash E. y Aruldoss M., 2021)	Modelos de detección de objetos	Examinan arquitecturas profundas, mencionando sus ventajas y desventajas. Se presentan conjuntos de datos y sus métricas de evaluación. También se mencionan ampliamente problemas en la detección de objetos.
(Abbas S., et al., 2022)	Modelos de detección de objetos	Se presentan una gran cantidad de modelos de aprendizaje profundo recientes, también se muestran conjuntos de datos y métricas de evaluación y se realiza una comparación de diversos modelos.
(Zou Z., et al., 2023)	Modelos de detección de objetos	Muestran la evolución de los algoritmos de detección de objetos durante la etapa de 1990-2019. Se presentan detectores fundamentales, conjuntos de datos, métricas, técnicas de aceleración, aplicación y desafíos y mejoras en los últimos años.

Tabla 1. Comparación de artículos de estudios sobre la detección de objetos [continuación].

3. Enfoques generales de detección de objetos

En esta sección se mencionarán brevemente algunos enfoques en la detección de objetos mediante técnicas clásicas y mediante técnicas de aprendizaje profundo. Con el paso del tiempo, los enfoques para la detección de objetos han ido cambiando debido a los avances que ha habido en los algoritmos del aprendizaje de máquina, sin embargo, los enfoques clásicos y mejoras a estos, siguen siendo utilizados. De igual manera, dado el incremento en las capacidades computacionales, en la cantidad de datos de entrenamiento, el desarrollo de algoritmos y de detectores profundos, se ha propiciado que las técnicas profundas sean ampliamente utilizadas en la actualidad.

3.1 Técnicas clásicas

Las técnicas clásicas son aquellas que no utilizan aprendizaje profundo. En esta sección se exponen ideas y métodos de este enfoque cuyas técnicas han predominado en la resolución de tareas de visión por computadora y, por ende, en la detección de objetos. Algunos ejemplos de estas técnicas se describen a continuación. *Color de pixel*: H. Shih y J. Chen (Shih H. y Chen J., 2021) consideran el color como característica para la detección de objetos. Utilizan el espacio de color HSI e YCbCr para detectar piel en imágenes de personas. De igual manera, A. Sáez et. al. (Saez A., et al., 2019) utilizan el color de un pixel y de sus vecinos para obtener la probabilidad de que un pixel pertenezca a cierta clase y así detectar lesiones en la piel mediante imágenes de dermatoscopia. *Template matching*: Es una técnica para detectar en una imagen, la región o regiones que más se asemejen a un patrón dado (Shou X., et al., 2019). *Características invariantes*: Son métodos más avanzados en la detección de objetos (Dash P. y Sigappi A., 2018; Ansari S., 2019; Gupta S., et al.), que extraen características invariantes ante el escalamiento, rotación, traslación como SIFT (Lowe D., 1999), SURF (Bay H., et al., 2006) y ORB (Rublee E., et al., 2011). *Patrones binarios locales (LBP)*: Son métodos basados en textura que han sido utilizados para la detección de rostros (Zhang B., et al., 2010; Zhang G., et al., 2004; Suruliandi A., et al., 2012; Sun N., et al., 2006). Ejemplo de este método y sus variantes es (Ojala T., et al., 1994). *Histograma de gradientes orientados (HOG)*: Es un método que se enfoca en la forma del objeto, y además brinda la información de la dirección de los gradientes. Algunas investigaciones que utilizan HOG para detectar objetos son (Seemanthini K. y Manjunath S., 2018; Nguyen N., et al., 2019; Li J., et al., 2019). *Los filtros de Haar*: Fueron propuestos por P. Viola y M. Jones en 2004 (Viola P. y Jones M., 2004). Han sido utilizados principalmente en la detección de rostros (Arfi A., et al., 2020; Rahmatulloh A., et al., 2021). En (Arunmozhi A. y Park J., 2018; Adouani A., et al., 2019) se publican comparaciones de HOG, LBP y características de Haar en la detección de vehículos y en la detección de rostros, respectivamente. *Filtros de correlación*: Son una clase de clasificadores, los cuales producen picos en la correlación cuando se localiza un objeto de interés. Estos filtros han sido utilizados en la detección y el seguimiento de objetos, como ASRCF (Adaptive spatial regularization correlation filter) (Dai K., et al., 2019). Para más información véase (Liu S., et al., 2021; Du S. y Wang S., 2022). En la Tabla 2 se resumen las citas de la información presentada en esta sección.

Artículo	Enfoque	Artículo	Enfoque
(Shih H. y Chen J., 2021; Saez A., et al., 2019)	Color del pixel	(Seemanthini K. y Manjunath S., 2018; Nguyen N., et al., 2019; Li J., et al., 2019)	Histograma de gradientes orientados
(Shou X., et al., 2019)	Template matching	(Arfi A., et al., 2020; Rahmatulloh A., et al., 2021)	Filtros de Haar
(Dash P. y Sigappi A., 2018; Ansari S., 2019; Gupta S., et al.)	Características invariantes	(Liu S., et al., 2021; Du S. y Wang S., 2022).	Filtros de correlación
(Zhang B., et al., 2010; Zhang G., et al., 2004; Suruliandi A., et al., 2012; Sun N., et al., 2006)	Patrones binarios locales		

Tabla 2. Algunos enfoques clásicos en la detección de objetos.

3.2 Técnicas profundas

Como se mencionó anteriormente, la detección de objetos tiene dos etapas: la localización y la detección. Dado que las redes neuronales convoluciones profundas (DCNN del inglés Deep Convolutional Neural Network) presentan un alto desempeño en tareas de clasificación, los métodos de detección de objetos más empleados en la actualidad están basados en DCNN. Este tipo de arquitecturas de detección de objetos puede dividirse en dos categorías: arquitecturas de detección de objetos de una etapa y de dos etapas (Zou Z., et. al. 2023).

3.2.1 Detectores de una etapa

Se caracterizan por localizar y clasificar el objeto en una sola etapa. Tienen la ventaja de presentar mayor velocidad de detección con respecto a los detectores de dos etapas, sin embargo, la precisión es menor que estos. A continuación, se hace una revisión de diversos detectores de una etapa.

En 2013 P. Sermanet *et al.* (Sermanet P., et al., 2013) mostraron OverFeat que consiste en utilizar la técnica multiscale and sliding window en la última capa de pooling de la DCNN para extraer las regiones de interés (donde puede haber un objeto) y predecir los valores de clasificación para cada región. En 2015 W. Liu et. al (Lu W., et al., 2016) propusieron la red SSD (Single Shot MultiBox Detector), cuya mayor contribución consistió en las técnicas de multirreferencia y multiresolución. La técnica de multirreferencia consiste en tener referencias (regiones rectangulares en distintos lugares en la imagen) y realizar las predicciones en dichas referencias; por otro lado, la técnica de multiresolución radica en hacer detecciones a diferentes escalas en diferentes capas de la red. J. Redmon et al. (Redmon J., et al., 2016) propusieron YOLO. Este detector divide la imagen en regiones y predice los rectángulos delimitadores y las probabilidades de que el objeto en dicha región pertenezca a cierta clase. J. Redmon y A. Farhadi mejoraron YOLO en (Redmon J. y Farhadr A., 2017) y (Redmon J. y Farhadr A., 2018). En (Redmon J. y Farhadr A., 2017) introdujeron la normalización por lotes para acelerar la convergencia y mejorar la generalización de la red. En (Redmon J. y Farhadr A., 2018) utilizaron la red FPN para que detectar más clases y objetos de múltiples tamaños. Más versiones de la red YOLO han sido presentadas (Terven J. y Cordova D., 2023), desde YOLOv4 hasta YOLOv8 (Bochkovskiy A., et al., 2020; Jocher G., 2020; Li C. et al., 2022; Wang C., et al., 2022; Jocher G., et al., 2023). En (Lin T., Goyal P., et. al., 2017) se tiene la red RetinaNet, donde se introduce una función de pérdida llamada «focal loss» para abordar el desequilibrio entre las clases del fondo y del primer plano. Esta función remodela la función de pérdida de entropía cruzada de modo que disminuye la pérdida asignada a muestras bien clasificadas. Por último, M. Hajizadeh et al. (Hajizadeh M., et al., 2023) presentan la red MobileDenseNet para mejorar la detección de objetos pequeños con un sistema embebido.

3.2.2 Detectores de dos etapas

Los detectores de dos etapas realizan la localización del objeto en una etapa y su clasificación en otra. En estos métodos primero se generan las regiones de propuesta (regiones candidatas a contener el objeto de interés) y después se extraen características de las regiones para determinar si el objeto se localiza en alguna de esas regiones. Su mayor ventaja con respecto a los detectores de una etapa es su mejor exactitud, sin embargo, suelen ser más lentos que ellos. Enseguida se describen algunos de los principales métodos de detección de objetos de 2 etapas presentados en la literatura.

R. Girshick et al. (Girshick R., et al., 2014) introdujeron los detectores de objetos en la era del aprendizaje profundo con la red Regions with CNN Features (RCNN), la cual combina algoritmos de generación de regiones de interés, seguido por la extracción de características con una CNN. Esta red tiene la importancia de haber introducido a los detectores de objetos en la era de las redes neuronales, pero tiene algunas desventajas como la generación de 2000 regiones de propuesta, lo cual consume espacio y recursos. K. He et al. (He K., et al., 2015) presentaron la red Spatial Pyramid Pooling Networks (SPPNet), cuya mayor contribución fue introducir la capa Spatial Pyramid Pooling (SPP), la cual le permite a la CNN generar un vector de características de tamaño fijo sin importar el tamaño de la imagen o región de interés. En el 2015, R. Girshick (Girshick R., 2015) mostró la red Fast RCNN. Fast RCNN entrena simultáneamente el detector y la regresión del rectángulo delimitador. En el 2017 S. Ren et al. (Ren S., et al., 2017) propusieron la red Faster RCNN, que incorpora una red de generación de propuestas (RPN). En la red Faster RCNN los bloques de detección de propuestas, extracción de características, regresión del rectángulo delimitador, etc, son integrados en un marco unificado. Después se desarrollaron otros métodos como R-FCN (Dai J., et al., 2016) y Light head RCNN (Li Z., et al., 2017). En 2017 T. Lin et al. (Lin T., Dollar P., et al., 2017) presentaron la Feature Pyramid Networks (FPN), donde su mayor contribución consiste en generar características de alto nivel semántico en todas las escalas, haciendo de esta una red apta para la detección de objetos pequeños. K. He et al. propusieron la red Mask R-CNN en el 2017 (He K., Gkioxari et al., 2017), esta red es una mejora de la red Faster R-CNN que permite también realizar una segmentación. En el 2019 G. Gkioxari et al. (Gkioxari G., et al., 2019) presentaron la red Mesh RCNN, la cual está basada en Mask RCNN. La red Mesh RCNN además de realizar la detección de objetos, también da una representación tridimensional de dichos objetos.

En la Tabla 3 y en la Figura 2 se muestra el resumen de citas de los detectores profundos presentados en esta sección.

Detectores de una etapa		Detectores de dos etapas	
Artículo	Red	Artículo	Red
(Sermanet P., et al., 2013)	OverFeat	(Girshick R., et al., 2014)	RCNN
(Lu W., et al., 2016)	SSD	(He K., et al., 2015)	SPPNet
(Redmon J., et al., 2016)	YOLO	(Girshick R., 2015)	Fast RCNN
(Redmon J. y Farhadr A., 2017)	YOLOv2	(Ren S., et al., 2017)	Faster RCNN
(Redmon J. y Farhadr A., 2018).	YOLOv3	(Dai J., et al., 2016)	R-FCN
(Bochkovskiy A., et al., 2020)	YOLOv4	(Li Z., et al., 2017)	Light head RCNN
(Jocher G., 2020)	YOLOv5	(Lin T., Dollar P., et al., 2017)	FPN
(Li C. et al., 2022)	YOLOv6	(He K., et al., 2017) ¡Error! No se encuentra el origen de la referencia.	Mask RCNN
(Wang C., et al., 2022)	YOLOv7	(Gkioxari G., et al., 2019)	Mesh RCNN
(Jocher G., et al., 2023)	YOLOv8		
(Lin T., Goyal P., et al., 2017)	RetinaNet		
(Hajizadeh M., et al., 2023)	MobileDenseNet		

Tabla 3. Enfoques con aprendizaje profundo para la detección de objetos.

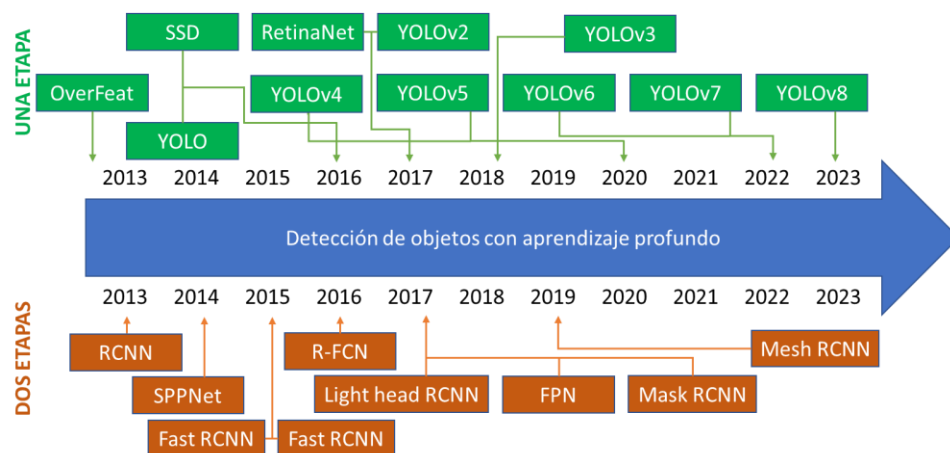


Figura 2. Línea de tiempo de arquitecturas de aprendizaje profundo para la detección de objetos.

4. Problemas en la detección de objetos

La detección de objetos es una tarea de visión por computadora que, a pesar de los avances en los últimos años, sigue presentando grandes dificultades. En esta sección se describen los problemas más comunes en la detección de objetos, y cómo han sido abordados recientemente en diversas investigaciones. Los problemas abordados en esta revisión son: oclusión, confusión, información contextual, cambios en la iluminación, objetos pequeños y cambios de escala, variación entre la misma clase y entre diferentes clases y deformación y cambios de pose.

4.1 Oclusión

Es un problema en el cual un objeto real o artificial oculta parcial o totalmente información de interés de una imagen.

La mayoría de los artículos analizados tienen como aplicación la detección de rostros y peatones ocluidos, tal situación se puede deber, principalmente, a un par de razones: 1) a la accesibilidad de los conjuntos de datos de entrenamiento cuyas clases son precisamente rostros o peatones, 2) y por el creciente interés en los vehículos autónomos donde es fundamental la correcta detección de personas, aunque estén ocluidas. El interés en los vehículos autónomos también se observa en (Liu Y., et al., 2021; Heo J., et al., 2022; Chilukuri D., et al., 2022), cuya aplicación consiste en detectar señales de tráfico y medios de transporte ocluidos. Por otra parte, existen metodologías en las cuales se analiza el mapa de características (Cen F., et al., 2022; Wang Y., et al. 2022) para comparar con dichos mapas regiones ocluidas o no ocluidas; o para que la red profunda se enfoque en regiones del mapa de características donde no hay oclusión. Otras investigaciones tienen como objetivo la detección de ciertas figuras geométricas (Zhao M., et al., 2021; Dong W., et al., 2021) y a partir de tal objetivo desarrollan una metodología específica. También existen métodos en los cuales se considera que los objetos que ocluyen son ruido, y se elimina ese ruido (Xiang L., et al., 2023). A continuación, se realiza un análisis de cada artículo mencionado en esta sección.

En (Moncef S. y Othman A., 2021) se abordan tres tipos de oclusiones para la detección de rostros: barba, bigote y lentes. Mediante el método de Fuzzy C-Means se extraen la barba y el bigote y con operaciones morfológicas se detecta la presencia de lentes. También W. Zheng et. al. (Zheng W., et al., 2020) utilizan información contextual para solucionar el problema de la oclusión. Este enfoque se inspira en las características de los nervios ópticos de los seres humanos para reconocer rostros con oclusión. Además, se hace un aumento de datos para obtener imágenes con oclusión y se propone aplicar few-shot para el reconocimiento. En (Zhou C. y Yuan J., 2020) se detectan peatones con oclusión mediante la correlación entre algunas partes del cuerpo humano. Dado que no se sabe cuáles partes estarán ocluidas, se aplica un conjunto de detectores de partes. Para una región candidata en la imagen, los K detectores darán K probabilidades, con lo cual se integran esas probabilidades para dar el resultado final indicando qué tan probable es que esa región contenga un peatón.

Para esto se utilizan tres métodos: Max integration, TopS integration y L1 integration. X. Zhou y L. Zhang (Zhou X. y Zhang L., 2022) crearon la red SA-FPN para detectar personas en imágenes de lugares concurridos, donde suele presentarse oclusión. Ellos diseñaron una red, Scale-FPN, para solucionar el problema de variaciones de escala y un mecanismo de atención (attention-based lateral connection) el cual mejora la información semántica y le permite al detector enfocarse en características relevantes de peatones con oclusión. X. Song et. al. (Song X., et al., 2020) presentan la red Progressive Refinement Network (PRNet) para la detección de peatones ocluidos. La red PRNet imita el proceso de anotación progresiva del ser humano en la detección de peatones oclusión; este proceso tiene tres etapas: 1) anotación de la parte visible (se analiza la región visible del peatón), 2) realización de un proceso estadístico (se estima si la región visible corresponde a un peatón) y, 3) una anotación del cuerpo completo (se estima dónde estaría la parte no visible). La PRNet realiza el proceso descrito con anterioridad al entrenarse de la siguiente manera: primero se entrena el módulo Visible-part Estimation (VE) para la detección de la parte superior del cuerpo humano, que corresponde a la anotación de la parte visible, luego el módulo Anchor Calibration (AC) que imita el proceso estadístico y por último el módulo Full-body Refinement (FR), el cual es como la anotación del cuerpo completo. VE y FR toman ground truth de partes visibles y de cuerpo completo como referencia, respectivamente. VE se entrena con el objetivo de aprender los rectángulos delimitadores A1 en las partes visibles del cuerpo humano; luego AC se encarga de migrar los rectángulos delimitadores A1 a rectángulos delimitadores de cuerpo completo A2. Una vez se tiene A2 se entrena FR. En (Song X., et al., 2022), se tiene la red PRNet++ (basada en (Song X., et al., 2020)) para mejorar la habilidad del modelo ante diferentes estados de oclusión (es decir, que el modelo se desempeñe bien en la detección de peatones con poca oclusión y mucha oclusión).

PRNet++ tiene dos ramas: Easy-branch, que se encarga de la detección con poca oclusión y Hard-branch, que se enfoca en la detección con mucha oclusión. Ambas ramas utilizan PRNet (Song X., et al., 2020). En (Wu C. y Ding J., 2018) se presenta un algoritmo, gradient direction-based hierarchical adaptive sparse and low-rank (GD-HASLR) para el reconocimiento de rostros ocluidos. Un método llamado Pose-Driven Visibility Model (PDVM) se propone en (Zhou S., et al., 2020) para la detección de personas. Ese método contiene tres partes: construcción de información de humanos no ocluidos, representación de características globales representativas y características alineadas en parte. R. Biswas et. al (Biswas R., et al., 2021) presentan el método de Estructura de Vecindario Dominante de Frecuencia One-Shot (OSD-DNS). En este método se obtiene la identidad de una cara ocluida mediante un clasificador entrenado con caras no ocluidas. H. Wang et. al. (Wang H., et al., 2021) proponen un modelo, la red Shrinking and Reweightint (SRNet) que puede aprender características globales más robustas y altamente discriminatorias, y obtener umbrales suaves a través del módulo DRS (Deep Residual Shrinkage Module) para eliminar las características relacionadas con el ruido para la detección de peatones.

Además, se diseña un modelo para la coincidencia de características parciales de puntos clave que indican si el objeto es un humano o no (RMPM): se asignan pesos adaptativos a cada característica parcial, las características parciales sin información útil (oclusión) se debilitan. El modelo SRNet utiliza como red pre-entrenada la ResNet50. En (Liu T., et al., 2020) se presenta un enfoque de detección de peatones con una red acoplada que está formada por subredes. Una de esas subredes se enfoca en el problema de la oclusión al utilizar una región de interés regional deformable, la cual genera características robustas de peatones con oclusión, ya que se adapta mejor a las variaciones de posición de partes de los peatones. Este enfoque utiliza la RFCN (Region-based Fully Convolutional Network) debido a que puede localizar un objeto aun cuando exista una información parcial del mismo.

En (Xie J., et al., 2020) se tiene un método para la detección de peatones con oclusión llamado mask-guided attention network (MGAN), el cual genera una máscara de atención espacial utilizando información de regiones visibles del cuerpo. Este método utiliza la red preentrenada Faster R-CNN. En (Wank K., et al., 2020) se incrementa el conjunto de entrenamiento al anotar imágenes con oclusión y diversidad de poses para la detección de rostros, además se propone la red Region Attention Networks (RAN), que captura de manera adaptativa la importancia de las regiones faciales, se utilizan las redes profundas pre-entrenadas ResNet-18 y VGG16. M. Flores et. al. (Flores M., et al., 2019) presentan un trabajo donde se divide la región de interés en 12 secciones, de esta manera se extraen las partes más relevantes; es decir, que tienen el potencial para identificar la forma humana. Para la extracción de características se usa el histograma de gradientes orientados (HOG) y para la clasificación se utilizan los algoritmos de máquina de vectores de soporte (SVM) e inferencia lógica (IL). J. Zuo et. al (Zuo J., et al., 2018) plantean un método con aprendizaje profundo para la detección de personas ocluidas, su método es Attention Framework of Person Body (AFPB), el cual consiste en dos componentes principales: un simulador de oclusión (OS), que genera muestras artificiales de personas ocluidas utilizando las muestras de personas sin ninguna oclusión; y un clasificador binario el cual es entrenado con las muestras sin oclusión y las muestras generadas con el OS. En (Zhou C. y Yuan J., 2020) se propone un enfoque que descubre patrones de oclusión que son representativos y discriminativos. Se basa en un modelo de partes deformables (DPM, del inglés Deformable Part Model) entrenado con muestras sin oclusión.

A partir de los patrones de oclusión, se buscan solamente los patrones que son representativas y que tienen un alto desempeño en la clasificación de manera individual. J. Dong et. al. (Dong J., et al., 2020) presentan un método para eliminar oclusiones basada en la red generativa adversaria (GAN). Este método consiste en dos GAN: G1, que representa un generador de oclusión, dada una imagen de un rostro y una imagen de oclusión, y G2 es la red que remueve la oclusión. Ambos generadores utilizan una arquitectura U-Net como codificador-decodificador. Y. Xu et. al. (Xu Y., et al., 2018) utilizan características profundas con forma de Omega (Ω), que son características de la región de cabeza y hombros de las personas para la detección de personas con oclusión. Se utiliza la red RFCN, la cual se entrena con muestras etiquetadas de todo el cuerpo humano, y otras imágenes etiquetadas de la parte de los hombros y la cabeza, haciendo el modelo más robusto ante posibles oclusiones. En (Wei W., et al., 2019) se utiliza la red profunda MobileNet para detectar peatones con oclusión mediante información de profundidad. C. Liu et. al (Liu C., et al., 2022) presentan la red Double Mask R-CNN para detectar personas en lugares concurridos, en donde existe oclusión. Esto lo logran al detectar puntos clave del cuerpo humano con la red MKFRCNN (Mask and Keypoint with Fast RCNN), de acuerdo con estos puntos clave se determina la visibilidad de las regiones donde están las personas. Si la visibilidad es menor que un cierto umbral, se añade una máscara para poder detectar las personas ocluidas.

Y. Liu et. al. en (Liu Y., Peng J., et al., 2021) proponen una red, TSingNet, para la detección de señales de tráfico ocluidas, la red aprende características sobre el contexto para la detección. Se utiliza la red ResNet50, y para obtener una información relevante del contexto, se reemplazan las capas de convolución de la última capa con tres bloques de fusión adaptativa de campo receptivo (ARFF). J. Heo et. al (Heo J., et al., 2022) crean la arquitectura Occlusion-aware Spatial Attention Transformer (OSAT), la cual está basada en Vision Transformer (ViT), CutMix augmentation y Occlusion Mask Predictor para resolver el problema de la oclusión. La utilización del ViT radica en implementar un mecanismo de atención para analizar medios de transporte con oclusión. En (Chilukuri D., et al., 2022) realizan una segmentación por instancias en imágenes de señales de tráfico con oclusión, posteriormente se realiza la detección. Gracias a la segmentación por instancias el modelo basado en la red SSD MobileNet atiende el problema de la oclusión.

En (Cen F., et al., 2021) se propone un enfoque para la clasificación de imágenes con oclusión al ajustar los parámetros de modelos pre entrenados con un conjunto de vectores de características profundas aumentado (DFVs). Los DFV son extraídos por una CNN. Se observa que el vector de diferencia (DV) entre los DFVs de una imagen ocluida y una sin oclusión está altamente relacionada con la oclusión. A partir de esto se propone un aumento de DFVs con pseudo-DFVs, que son generados de manera aleatoria al añadir DVs extraídos de un conjunto pequeño de pares de imágenes limpias y con oclusión. En (Li F., et al., 2022) se propone la red Generative Adversarial Occlusion Network (GAON), la cual se entrena junto con la red Faster R-CNN. Mientras la Faster R-CNN aprende a reconocer imágenes de varias clases, la GAON aprende a generar oclusiones en el mapa de características de las capas escondidas de la Faster R-CNN; de esta manera el detector es robusto ante la oclusión. En (Wang Y., et al., 2022) se presenta una red llamada OD-UTDNet. Esta red tiene un módulo (Dilated Attention Cross Stage Partial) el cual es responsable de dos aspectos principales: 1) hace que la red OD-UTDNet se enfoque en las regiones sin oclusión y 2) mejora la capacidad de la red OD-UTDNet para extraer características en regiones con oclusión.

Una técnica basada en triángulos inscritos para la detección de círculos ocluidos se presenta con M. Zhao et. al. (Zhao M., et al., 2021), en (Dong W., et al., 2021) se detectan formas elípticas con oclusiones muy grandes mediante una modificación de la red R-CNN: Ellipse R-CNN, para manejar mejor la oclusión se integró la red U-Net para generar mapas de características decodificados que contienen información escondida. Además, se aprenden varios patrones de oclusiones.

L. Xiang et. al. (Xiang L., et al., 2023) consideran que en una región de interés la información que ocluye es ruido, a partir de esto diseñan un proceso iterativo llamado DINF (Dynamic instance noise filter) para mejorar la relación señal-ruido de las características de los objetos de interés. DINF primeramente transforma las características obtenidas en las regiones de interés, RDI, a un dominio en el cual los valores de la señal de ruido son cercanos a cero. Después, se utiliza una umbralización para remover el ruido. En la Tabla se hace un resumen de los artículos previos.

Artículo	Método	Artículo	Método
(Moncef S. y Othman A., 2021)	Con Fuzzy C-Means se extrae la barba y el bigote, y mediante operaciones morfológicas se detectan los lentes para la detección de rostros.	(Zuo J., et al., 2018)	Generación de muestras con oclusión a partir de muestras sin oclusión.
(Zheng W., et al., 2020)	Con la información contextual y el aumento de datos se detectan objetos con oclusión.	(Zhou C. y Yuan J., 2020)	Se basa en un modelo de partes deformables.
(Zhou C. y Yuan J., 2020)	Al utilizar un detector de partes humanas se detectan peatones con oclusión.	(Dong J., et al., 2020)	Dos redes generativas adversarias (GAN), una de ellas genera oclusión y la otra la elimina.

Tabla 4. Resumen de los artículos que atienden el problema de oclusión.

Artículo	Método	Artículo	Método
(Zhou X. y Zhang L., 2022)	Un mecanismo de atención mejora la información semántica, gracias a esto el detector se enfoca en características de peatones con oclusión.	(Xu Y., et al., 2018)	Con características de la región de la cabeza y hombros de humanos (características omegas) se hace más robusta la red R-FCN ante los objetos con oclusión.
(Song X., et al., 2020)	Inspirado en el proceso de anotación utilizado por el ser humano para la detección de personas con oclusión el cual consiste en: analizar la región visible, se estima si la región es de una persona y si sí se estima la región completa de la persona.	(Wei W., et al., 2019)	Con el uso de información de profundidad se resuelve el problema de la oclusión.
(Song X., et al., 2022)	La red PRNet++ tiene dos ramas, una rama se hace experta en la detección con poca oclusión y la otra se hace experta en la detección con alta oclusión.	(Liu C., et al., 2022)	Se detectan puntos clave del cuerpo humano para determinar la visibilidad de personas ocluidas.
(Wu C. y Ding J., 2018)	Con base al método <i>gradient direction-based hierarchical adaptive sparse</i> se detectan objetos con oclusión.	(Liu Y., Peng J., et al., 2021)	La red TSingNet aprende características contextuales para la detección de señales de tráfico con oclusión.
(Zhou S., et al., 2020)	A partir de características globales y la información en imágenes de personas sin oclusión se detectan personas con oclusión.	(Heo J., et al., 2022)	Implementación de un mecanismo de atención para analizar los objetos con oclusión.
(Biswas R., et al., 2021)	Se utiliza OSD-DNS para reconocer caras ocluidas.	(Chilukuri D., et al., 2022)	Segmentación por instancias.
(Wang H., et al., 2021)	Asignación de pesos adaptativos para debilitar las características con oclusión.	(Li F., et al., 2022)	Se presenta una red adversaria para generar oclusiones en el mapa de características de las capas más profundas de la red Faster R-CNN.
(Liu T., et al., 2020)	Se tienen subredes acopladas. Una de esas subredes utiliza una región de interés deformable, la cual extrae características robustas de peatones con oclusión.	(Wang Y., et al., 2022)	A partir de regiones sin oclusión se mejora la capacidad para extraer características en regiones con oclusión.
(Xie J., et al., 2020)	La red <i>mask guided attention network</i> se enfoca en regiones visibles del cuerpo humano.	(Zhao M., et al., 2021)	Mediante triángulos inscritos se detectan círculos ocluidos.

(Wang K., et al., 2020) y (Cen F., et al., 2021)	Se utiliza la diferencia de vectores con características en imágenes con oclusión y sin oclusión.	(Dong W., et al., 2021)	Con la integración de las redes Ellipse R-CNN y U-Net se detectan formas elípticas.
(Flores M., et al., 2019)	La imagen se divide en secciones para extraer las regiones que tienen el potencial de ser partes humanas. Las características se extraen con HOG.	(Xiang L., et al., 2023)	Se considera que la información en una región de interés se divide en dos: objeto de interés y ruido, donde el ruido es la información de los objetos que ocuyen. A partir de esto, se diseña un modelo para eliminar este ruido.

Tabla 4. Resumen de los artículos que atienden el problema de oclusión [continuación].

4.2 Confusión

La confusión, o cluttering, es un problema de detección de objetos que se presenta cuando el fondo de la imagen contiene color o textura semejantes al del objeto de interés. En la literatura existen varios trabajos desarrollados recientemente para lograr la detección de objetos en este tipo de escenarios.

La metodología más utilizada para resolver este problema consiste en diseñar arquitecturas para encontrar y mejorar los bordes de los objetos de interés (Skoviera R., et al., 2018; Xu X., et al., 2021; Ji G., et al., 2022; Chen T., et al., 2022; Zhang C., et al., 2022; Zhuge M., et al., 2022). La razón de esto se debe a la naturaleza misma del problema: el contorno (bordes) de los objetos que se confunden con el fondo (objetos camuflados) es difícil de identificar. Con lo cual, si se logra obtener correctamente los bordes de los objetos camuflados, la detección resultará más sencilla. Existen otras metodologías menos utilizadas basadas en información contextual, enfoque en primer plano, etc. (Liu S., et al., 2021; Huang X., et al., 2021; Zhang T., et al., 2020; Zhou S., et al., 2019; Wang K., et al., 2021; Kreim A., et al.; Chen G., et al., 2022). Sin embargo, se aprecia que la tendencia para la detección de objetos camuflados está en la detección de bordes. Tal como la oclusión, la detección de personas y vehículos también es una aplicación de interés en la detección de objetos camuflados (Zheng Y., et al., 2019; Wang Y., et al., 2020; Wei X., et al., 2020). A continuación, se presentan a más detalle estos artículos.

R. Skoviera et. al. (Skoviera R., et al., 2018) proponen el uso de una red bio inspirada llamada Hierarchical Temporal Memory (HTM) para la detección de múltiples objetos en imágenes con confusión. El objetivo de la HTM es aprender a detectar representaciones invariantes de la entrada. Para mejorar el desempeño de la detección basado en HTM, se propone un sistema que usa información de tres HTM que se enfocan en diferentes características de la imagen: ejes, textura y color. El conjunto de datos fue generado para representar el cluttering. X. Xu et. al. (Xu X., et al., 2021) detectan el contorno de objetos en imágenes complejas mediante la red Boundary Guidance Network, la cual consiste en el ILD (Initial Localization Decoder), para encontrar la posición inicial de los objetos camuflados lo más exacto posible, y en el RRD (Residual Refinement Decoder), para refinar los detalles de los bordes de los objetos. En (Ji G., et al., 2022) se presenta la red ERRNet (Edge-based reversible re-calibration network) para la detección de objetos con confusión. Esta red presenta dos módulos: Selective Edge Aggregation y Reversible Re-calibration Unit para que la red logre una detección efectiva de los bordes y una comparación entre regiones potenciales camufladas y el fondo. T. Chen et. al. (Chen T., et al., 2022) proponen una red llamada Boundary-guided Network (BgNet) que tiene el módulo Locating Module (LM) que se compone de una rama para obtener información contextual y otra rama que se utiliza para captar detalles de bajo nivel.

Las características de alto (información contextual) y bajo nivel se combinan, así si la información contextual es deficiente, las características de bajo nivel pueden dar información para lograr una correcta detección. También está el módulo Boundary-guided Fusion Module (BFM), el cual toma tres entradas: el mapa de características obtenido con LM y dos mapas de predicciones, uno que es la máscara del objeto y el otro la de los bordes; con estas entradas se hace un mejoramiento de características. Este mejoramiento tiene dos pasos, en el primero se calculan las características del fondo y del plano de interés y en el segundo se obtiene el resultado final. C. Zhang et. al. (Zhang C., et al., 2022) implementan dos módulos para la detección de objetos camuflajeados. Un módulo es el Neighbor Connection Mode (NCM) el cual concatena características de la capa previa, de la capa actual y de la capa siguiente; a estas características concatenadas se les llama características intermedias después de la fusión (CIF).

El otro módulo es el Hierarchical Information Transfer (HIT), cuya entrada son las CIF y su arquitectura consiste en 5 ramas: cuatro de ellas son convoluciones, suma y multiplicación de elementos, función sigmoide y al final de estas cuatro ramas se concatenan, y la otra rama implementa un aprendizaje residual y un mecanismo de atención para omitir características irrelevantes. El módulo NCM resalta los bordes y la información de posición de los objetos detectados mientras que el módulo HIT extrae características, con la combinación de ambos se realiza la detección. M. Zhuge et. al. (Zhuge M., et al., 2022) presentan la red CubeNet, que utiliza características jerárquicas de múltiples capas para la detección de objetos camuflajeados. La red CubeNet consiste en dos square fusión decoder (SFD) y en un sub edge decoder (SED). El SFD utiliza características de alto y bajo nivel extraídas por bloques encoder-decoder. El SED está localizado entre cada SFD para mejorar la representación de los bordes.

S. Liu et. al (Liu S., Liu D., et al., 2021) proponen un mecanismo para la actualización de plantillas para mejorar el seguimiento de objetos en fondos complejos. Este mecanismo utiliza múltiples plantillas para estimar la ubicación de los objetos. X. Huang et. al (Huang X., et al., 2021) presentan una red neuronal bio inspirada basada en un filtrado temporal neurodinámico y un filtrado espacial de Gabor. T. Zhang et. al (Zhang T., et al., 2020) proponen un enfoque que consiste en incrementar el conjunto de entrenamiento a partir de muestras no etiquetadas para localizar y clasificar animales salvajes en imágenes.

Dicho enfoque se basa en tres pasos: 1) entrenar un modelo (llamado modelo maestro) con muestras etiquetadas manualmente, como en el aprendizaje supervisado, 2) se ejecuta el modelo maestro sobre muestras no etiquetadas para seleccionar más muestras y 3) se entrena un modelo (llamado modelo estudiante) con las muestras etiquetadas manualmente y las muestras seleccionadas en el paso 2. S. Zhou et. al (Zhou S., et al., 2019) exponen una red neuronal profunda para la detección de personas camuflajeadas. La red aprende características discriminativas del primer plano de cada imagen. Una red neuronal llamada Foreground Attention Neural Network (FANN), mejora el primer plano y debilita el fondo.

Una subred guía la atención de la red, donde un codificador es utilizado para reconstruir una máscara binaria, y un decodificador enfoca su atención en las personas en el primer plano. K. Wang et. al (Wang K., et al., 2021) detectan objetos camuflajeados a partir de cómo el mecanismo visual humana lo hace: normalmente es difícil encontrar los objetos camuflajeados en una primera observación. Para eso se pasa a una segunda etapa, donde se realiza un análisis de la primera etapa. Con esto en mente se propone el modelo D2C-Net, que contiene dos módulos: Dual-branch Features Extraction (DFE) y Gradually Refined Cross Fusion (GRCF). El módulo DFE simula el mecanismo de dos etapas del sistema de visión del ser humano: primero se realiza una concatenación densa para agregar características de varios niveles para expandir el campo receptivo.

En el módulo GRCF se combinan características para mejorar el rendimiento de la detección. En (Kreim A., et al.) se propone un conjunto de datos llamado PTAW217Synth para la generación de muestras sintéticas para el seguimiento de personas en ambientes complicados (lluviosos, nevados, con mucha neblina), para mejorar el desempeño de algoritmos profundos de seguimiento de personas. En (Chen G., et al., 2022) se tiene la red Context-aware Cross-level Fusion Network (C2F-Net), dicha red fusiona características del contexto y características de nivel cruzado (esto es, a partir de las características más profundas se refinan las características de bajo nivel) para la detección de objetos camuflajeados.

En (Zheng Y., et al., 2019) se utiliza una red convolucional para detectar personas con camuflaje. Inicialmente se construye el conjunto de datos de gente camuflajeadada en escenas naturales. Luego, con una red neuronal convolucional basada en la red VGG16, se extraen características semánticas y se introducen conexiones en la fase de deconvolución para lidiar con la influencia de patrones de camuflaje y el ruido en el fondo. Finalmente, se realiza una segmentación con super píxeles y se aplican restricciones espaciales de suavizado para mejorar la detección. En (Wang Y., et al., 2020) se detectan tanques en un entorno complejo (con mucha confusión entre el objeto y el fondo) con el detector YOLO utilizando pocas imágenes de entrenamiento. X. Wei et. al presentan en (Wei X., et al., 2020) el modelo profundo PftNet para la detección de personas en minas. Este modelo está basado en una red de transferencia de características en paralelo y está compuesto por dos módulos interconectados: el módulo de identificación, para ajustar la ubicación y el tamaño de los rectángulos delimitadores, y de localización, para que el modelo se adapte a varias escalas y relaciones de aspecto. En la Tabla 5 se comparte un resumen de los artículos revisados en esta subsección.

Artículo	Método	Artículo	Método
(Skoviera R., et al., 2018)	Red bioinspirada llamada Memoria Temporal Jerárquica (HTM). Se utiliza información de 3 HTM que se enfocan en características de ejes, textura y color. Además, el conjunto de datos fue generado para representa <i>cluttering</i> .	(Zhang T., et al., 2020)	Se utiliza la red Faster R-CNN y se consiguen rectángulos delimitadores a los cuales se les aplican transformaciones aleatorias.
(Xu X., et al., 2021)	Detección de contornos.	(Zhou S., et al., 2019)	Una red neuronal se enfoca en el primer plano y debilita el fondo para la detección de personas.
(Ji G., et al., 2022)	Detección de bordes y una comparación entre los objetos y el fondo.	(Wang K., et al., 2021)	Inspirado en la visión humana.
(Chen T., et al., 2022)	Con dos módulos: LM, para obtener información contextual e información de bajo nivel; y BFM para calcular las características del fondo y del objeto de interés.	(Kreim A., et al.)	Conjunto de datos para la detección de personas en ambientes complejos.
(Zhang C., et al., 2022)	Mediante los módulos NCM e HIT, para resaltar los bordes y la información de la posición de los objetos y la extracción de características, respectivamente.	(Chen G., et al., 2022)	A partir de características más profundas se mejoran las características de bajo nivel.
(Zhuge M., et al., 2022)	Bloques <i>encoder-decoder</i> extraen características de alto y bajo nivel, además se mejora la información de los bordes.	(Zheng Y., et al., 2019)	Con características semánticas y conexiones en la etapa de deconvolución para atender el camuflaje y el fondo.
(Liu S., Liu D., et al., 2021)	Múltiples <i>templates</i> .	(Wang Y., et al., 2020)	YOLO.
(Huang X., et al., 2021)	Red bioinspirada con un filtrado temporal neurodinámico y filtro de Gabor.	(Wei X., et al., 2020)	Transferencia de características.

Tabla 5. Resumen de los artículos que atienden el problema de confusión.

4.3 Información contextual

La información contextual se refiere a que el sistema de visión utilice la relación entre objetos de la imagen para poder comprender mejor la escena y detectar objetos. Con lo cual, el cómo adquirir información contextual en una imagen para poder localizar y clasificar objetos es un problema de interés en el área de detección de objetos.

Tras realizar un análisis de los artículos mostrados en la Tabla 6 en los cuales se presentan metodologías para la detección de objetos utilizando información contextual se encontró que todos siguen una metodología similar. Esto no es de extrañar, pues la información contextual se basa en encontrar la relación entre objetos de la imagen. Por eso, las metodologías consultadas tienen como idea fundamental el encontrar relaciones en la imagen. Estas relaciones se encuentran de diversas maneras: extrayendo características locales y globales, combinando las características locales y globales, utilizando características con más sentido semántico, proporcionando información de píxeles vecinos, usando objetos que se detecten con alta precisión para detectar otros objetos. Sin embargo, resulta que la extracción de características globales y locales y su combinación es la línea a para seguir para la obtención de información contextual. Los artículos revisados se describen a continuación.

En (She X., et al., 2021) se presenta una red profunda híbrida, ScieNet, con extracción de información contextual asistida por una red spike (SNN). ScieNet tiene una SNN con un algoritmo Spiking-timing-dependent plasticity (STDP) que extrae la información contextual. Luego se tiene una red profunda para la clasificación cuya entrada es la información contextual obtenida con STDP. S. Xie et. al. (Xie S., et al., 2020) presentan un método que integra características de diferentes campos receptivos para obtener una información contextual, mediante una red llamada diverse receptive field network (DRFNet). La utilización de varios módulos de campos receptivos en la DRFNet hace que regiones discriminativas en la vecindad del objeto de interés obtengan respuestas satisfactorias para la detección.

Esto indica que los módulos receptivos obtienen información contextual y detalles locales, con lo cual el detector puede distinguir mejor el fondo del objeto durante el entrenamiento de la red. X. Liang et. al (Liang X., et al., 2020) proponen el detector Feature Fusion and Scaling-based single shot detector (FS-SSD) para la detección de objetos pequeños con información contextual en imágenes aéreas. Además de las características profundas aprendidas por el FS-SSD, se mejora la precisión de la detección utilizando contexto espacial. La idea general del análisis del contexto espacial es utilizar los objetos detectados con un alto grado de precisión para objetos menos fiables. Q. Zhang et. al (Zhang Q., et al., 2020) muestran la red Dense Attention Fluid Network (DAFNet) para la detección de objetos sobresalientes en una imagen utilizando información contextual. DAFNet es una arquitectura encoder-decoder (basada en la red VGG16) donde en cada bloque de convolución se tiene un módulo llamado Global Context-aware Attention (GCA). El módulo GCA consiste en dos componentes: Global Feature Aggregation (GFA) y Cascaded Pyramid Attention (CPA). El módulo GFA toma las características obtenidas por la red implementada y produce características que codifican la información contextual global; este módulo tiene como objetivo lograr que las características se alineen y que exista un refuerzo mutuo entre los patrones sobresalientes al agregar relaciones semánticas globales entre pares de píxeles.

El módulo CPA se utiliza para abordar la variación en escala de los objetos; tiene como entrada las características obtenidas con el módulo GFA, el módulo CPA produce un mapa de atención refinado progresivamente. S. Wang et. al. (Wang S., et al., 2020) proponen un método que imita la percepción visual humana para la detección de navíos. En ese trabajo se desarrollan tres señales visuales para modelar el contexto de las imágenes: señal de escasez, debido a que los navíos solamente ocupan una pequeña porción de la escena; señal de contraste, ya que los navíos se diferencian notablemente del fondo; y señal de concentración, que se enfoca en la estructura local de los navíos. Una vez se tienen esas tres señales visuales, se combinan en un mapa de visibilidad para proveer ayudas visuales para la detección. En (Han L., et al., 2021) se propone la red Context and Structure Mining Network (CSMN) para la detección de objetos en video usando información del contexto. La CSMN consiste en un módulo de codificación de la información contextual espaciotemporal (stCIE) y un módulo de agregación de características de objetos basado en la estructura. La idea de stCIE consiste en que a cada mapa de características (obtenido con la red ResNet-101) de un pixel del objeto de interés se le agregan los píxeles vecinos tanto en el espacio como en el tiempo. En (Zhang L., et al., 2020) se propone aumentar una red neuronal feedforward utilizando un mecanismo de refinamiento en varias etapas.

En la primera etapa se construye una red maestra para generar un mapa de predicción en el cual faltan las estructuras más detalladas. En las etapas siguientes, la red de refinamiento con conexiones recurrentes a la red maestra combina información de contexto local a través de las etapas para refinar el mapa anterior. La información contextual propuesta se basa en que las redes neuronales en las capas más profundas tienen una fuerte consistencia semántica debido al largo campo receptivo, mientras que en las capas menos profundas las características codifican estructuras y detalles con poca consistencia semántica debido a su pequeño campo receptivo.

Para aprovechar lo antes mencionado, se aplica el módulo Channel Attention Mechanism (CAM), donde las características de alto nivel proveen una guía para extraer características discriminativas en las capas más bajas. También se implementa una capa global average pooling (GAP) en la cima de la red maestra para reunir información contextual global con un campo receptivo largo e introduce consistencia semántica en los siguientes módulos CAM. H. Luo et. al. (Luo H., et al., 2021) proponen el método Feature Fusion Semantic Information Enhancement (FFSI). Este método tiene dos módulos: el módulo de mejoramiento de información contextual (CIE, Context Information Enhancement) y el módulo de mejoramiento del campo receptivo (RFE, Receptive Field Enhancement).

El CIE resalta la ubicación de los objetos al establecer una relación entre la información contextual local y global. Mientras que el RFE mejora el campo receptivo al utilizar convolución dilatada para adaptar la detección de objetos de diferentes escalas, especialmente objetos pequeños. J. Guo et. al (Guo J., et al., 2020) se propone un detector de objetos con información contextual (EGCI-Net), el cual es una red de detección que tiene dos partes: una red backbone para extraer características y un modelo piramidal para enriquecer la información de contexto global de forma jerárquica. La información de contexto se obtiene con el módulo Pyramid Feature Pool (PFPM), que consiste en dos partes: el módulo pyramid pooling (PPM) y el módulo multi-scale feature pool (MFP). El PPM genera una priorización contextual mediante la fusión de un pooling de promedios multiescala, y el MFP genera información contextual piramidal mediante el uso de un pooling multiescala en el contexto global. Y. Zhu et. al (Zhu Y., et al., 2019) muestran la red Attention CoupleNet (ACoupleNet). Esta red utiliza la red ResNet-101 para la extracción de características. La red ACoupleNet funciona de la siguiente manera. Dada una imagen, se extraen características de los objetos utilizando atención en cascada.

Luego, con una red de regiones de propuesta (RPN) se generan las regiones donde puede estar un objeto. Cada propuesta va a dos ramas FCN (Fully Connected Networks) distintas: FCN local y FCN global, para extraer la información de la estructura global y aprender las partes específicas de objetos, respectivamente. Por último, en la salida se acoplan las dos ramas para predecir la clase del objeto. La FCN local captura diversas partes específicas (por ejemplo, en un rostro humano detecta la nariz, boca, etc., las cuales corresponden con regiones con respuesta alta en el mapa de características), lo cual refleja de manera efectiva las propiedades locales del objeto visual (siendo útil para lidiar con la oclusión o cuando el contorno del objeto está incompleto), mientras que la FCN global describe al objeto utilizando características de la región completa. Y. Gong et. al (Gong Y., et al., 2020) presentan la red Context Aware (CA-CNN) para la detección de objetos (la red neuronal convolucional es una Faster RCNN basada en VGG 16). La CA-CNN está constituida por la generación de propuestas, extracción de características de contexto, fusión de características y clasificación. Para extraer la información de contexto se propone una capa de extracción de regiones de interés, RDI, con contexto.

La RPN (red de generación de propuestas) conserva solamente las 256 mejores propuestas como RDI, cuyo tamaño es exactamente el de los objetos reales, por esto se generan las RDI con contexto para extraer información de contexto. G. Zhang et. al (Zhang G., et al., 2019) presenta una red, CAD-Net, para la detección de objetos con conciencia del contexto. La CAD-Net consiste en: 1) una red de contexto global (GCNet, Global Context Net) que aprende la correlación entre los objetos de interés y su correspondiente escena global, es decir, la correlación entre las características de los objetos y las características de la imagen total (la GCNet utiliza la red ResNet-101) y 2) la red Pyramid Local Context (PLCNet) que aprende la coocurrencia de características multiescala y/o la coocurrencia de objetos alrededor del objeto de interés (PLCNet utiliza Fast RCNN). W. Zhang et. al (Zhang W., et al., 2021) proponen un modelo llamado global context aware (GCA) para fortalecer la correlación espacial entre el fondo y el plano de interés fusionando la información de contexto global.

El módulo que se encarga de la información del contexto consiste en dos submódulos: módulo de atención y módulo de desvinculación de tareas. El módulo de atención se encarga de mapear la información de contexto global a un espacio de mayor dimensión y el módulo de desvinculación da el resultado de la clasificación y posicionamiento. En (Yuan Y., et al., 2019) se propone un método para la detección de señales de tráfico en ambientes complejos (se utiliza la red MobileNet como backbone).

El proceso principal del método propuesto consiste en dos componentes: el módulo de aprendizaje de características multiresolución, el cual combina diferentes niveles características semánticas con capas de deconvolución densamente conectadas; el otro módulo es el Vertical Spatial Sequence Attention (VSSA), que codifica la información vertical para una clasificación más exacta de las señales de tráfico. En (Siris A., et al., 2021) se propone un enfoque de aprendizaje con conciencia del contexto utilizando la red Mask-RCNN como backbone.

Específicamente, se proponen dos módulos: el módulo Semantic Scene Context Refinement (SSCR), para mejorar las características contextuales de los objetos sobresalientes de la imagen y el módulo Contextual Instance Transformer (CIT) para aprender las relaciones contextuales entre los objetos y la escena. H. Xie et. al (Xie H., et al., 2019) presentan un método de detección de peatones, Deconvolution Integrated Faster R-CNN (DIF R-CNN), que consta de tres componentes: generación inicial de mapas de características (que utiliza la red Inception-ResNet), módulo de deconvolución, que genera un mapa de características sintético y aporta información contextual, y por último el módulo de detección y región de propuesta. K. Zhang et. al. (Zhang K., et al., 2021) proponen una red semántica con conciencia del contexto (SCANet) para la detección de objetos en multiescala. Se diseñan dos módulos: Receptive Field-Enhancement (RFEM), que extrae características a diferente escala, y el módulo Semantic Context Fusion (SCFM), que combina características de alto nivel semántico con características de bajo nivel. En (Shi P., et al., 2023) proponen el detector Swin Deformable Transformer-BiPAFPN-YOLOX.

Este detector tiene un mecanismo de atención llamado Reconstructed Deformable Self-Attention, el cual guía, mediante las regiones importantes en el mapa de características, las relaciones existentes entre parches de la imagen (así obtiene información contextual). Además, se tiene una red, BiPAFPN, para fusionar características multiescala y, por último, se utiliza la red YOLOX para realizar la clasificación y regresión usando las características obtenidas previamente. En la Tabla 6 se condensa la información previamente mencionada.

Artículo	Método	Artículo	Método
(She X., et al., 2021)	Red ScieNet que extrae información contextual y es asistida por una red <i>spike</i> .	(Gong Y., et al., 2020)	Extracción de regiones de interés con contexto.
(Xie S., et al., 2020)	Integración de distintas características de diferentes campos receptivos.	(Zhang G., et al., 2019)	Correlación entre los objetos de interés y la escena global.
(Liang X., et al., 2020)	Los objetos detectados con alta precisión se utilizan para detectar objetos menos fiables.	(Zhang W., et al., 2021)	Modelo <i>global context aware</i> que fusiona la información de contexto global.
(Zhang Q., et al., 2020)	Módulo consciente del contexto compuesto por una adición de características y atención en cascada piramidal.	(Yuan Y., et al., 2019)	El módulo de secuencia de atención espacial vertical para codifica la información vertical para la detección de señales de tráfico.
(Wang S., et al., 2020)	Inspirada en la visión humana para la detección de navíos.	(Siris A., et al., 2021)	Mejoramiento de las características contextuales y aprendizaje de las relaciones contextuales entre los objetos y la escena.
(Han L., et al., 2021)	A cada mapa de características de un pixel se le añaden los pixeles vecinos.	(Xie H., et al., 2019)	Generación de características de contexto con un módulo de deconvolución.
(Zhang L., et al., 2020; Luo H., et al., 2021)	Las características de alto nivel semántico son una guía para la extracción de características en las capas más bajas.	(Zhang K., et al., 2021)	Se tiene un módulo para mejorar el campo receptivo y otro que fusiona la información semántica.
(Guo J., et al., 2020)	Información del contexto con un módulo de características piramidales.	(Shi P., et al., 2023)	Se encuentran relaciones que existen entre parches de la imagen.
(Zhu Y., et al., 2019)	Una FCN local que captura diversas partes específicas y una FCN global que describe al objeto utilizando las características de toda la región.		

Tabla 6. Resumen de los artículos que atienden el problema de información contextual.

4.4 Cambios en la iluminación

El problema en el cambio en la iluminación consiste en que un objeto puede verse completamente diferente ante variaciones en la iluminación de una escena. Dado que la visión por computadora lidia con imágenes naturales, donde la iluminación es variable, es fundamental que el sistema de visión sea robusto ante los cambios en la iluminación. Algunos han demostrado que incrementar el conjunto de entrenamiento, de tal modo que los objetos de interés estén con diferente iluminación, hace más robusto al sistema de visión ante cambios de iluminación (Wu D., et al., 2020; Olarewaju M., 2021; Badeka E., et al., 2020; Li G., et al., 2020; Wu Y., et al., 2020), siendo esta la metodología más seguida, pues el modelo de sistema de visión es capaz de aprender a detectar objetos ante distintas iluminaciones. En otras investigaciones se ha propuesto mejorar la iluminación de las imágenes (Xiao B., et al., 2021; Zhou K., et al., 2020) o utilizar características invariantes a la iluminación (Mohanty S., et al., 2019) y características manuales (Kumar S., et al., 2021). También se ha explorado el uso de imágenes infrarrojas (Dai X., et al., 2021). En otros casos se encuentra que ante ciertos conjuntos de datos hay redes profundas que son robustas al cambio de iluminación (Afif M., et al., 2020). A continuación, se hace una breve descripción de los trabajos antes mencionados.

D. Wu et. al (Wu D., et al., 2020) utilizan la red YOLO v4 para detectar flores. El conjunto de entrenamiento con el cual se ajustaron los parámetros de la red contiene imágenes tanto de flores con iluminación frontal como trasera, haciendo que la red pueda clasificar flores correctamente ante varios cambios de iluminación. M. Olarewaju (Olawaju M., 2021) modifica la red YOLOv3, para proponer la red YOLO-Tomato-A, la cual detecta tomates en imágenes naturales. Todas las imágenes del conjunto de muestras fueron obtenidas en ambientes y condiciones naturales: tomates con oclusión, traslape y distintas variaciones de iluminación. Este hecho hace que la red sea robusta ante la detección de tomates en imágenes con distinta iluminación. Un método similar se sigue en (Badeka E., et al., 2020), donde se entrena la red YOLOv3 con imágenes obtenidas en ambientes complicados, con oclusión, variaciones de iluminación, etc. En (Li G., et al., 2020) se utiliza la red Faster R-CNN para la detección de peatones en imágenes obtenidas por un dron. El conjunto de muestras para el entrenamiento contiene 1500 imágenes con varias condiciones climáticas, imágenes de día y de noche, haciendo el modelo más robusto ante cambios de iluminación. Asimismo, en (Wu Y., et al., 2020) se hace un aumento de datos en el conjunto de entrenamiento para la detección de señales de tráfico en distintas horas del día.

B. Xiao et. al. (Xiao B., et al., 2021) proponen un método para el seguimiento de máquinas de construcción en la noche. El método consta de cinco módulos: de mejoramiento de la iluminación, detección, seguimiento con filtros Kalman, asociación y asignación líneas. El módulo de mejoramiento de la iluminación es en el que se resuelve el problema de la variación de la iluminación en la detección de objetos, con lo cual nos enfocaremos únicamente en ese módulo. En este módulo se utiliza la red GLADNet (global illumination-aware and detail-preserving network) para la mejora de la iluminación en imágenes oscuras preservando la mayoría de los detalles. Esta red está dividida en dos etapas: estimación de la distribución de iluminación y la reconstrucción de detalles. En la primera etapa un mapa de características pasa por una red encoder-decoder para estimar la iluminación global en la imagen. La red encoder utiliza una CNN para submuestrear las características, mientras que la red decoder (que utiliza una CNN también) las sobremuestra. La etapa de reconstrucción de detalles concatena los mapas de características de la primera etapa con los de la imagen de entrada. Luego, este mapa de características concatenado es procesado por tres capas convolucionales para preservar una mayor cantidad de detalles de la imagen de entrada.

La salida del módulo de mejoramiento de la iluminación es la entrada del módulo de detección. En (Zhou K., et al., 2020) se propone la red Modality Balance Network, MBNet, para la detección de peatones. La red MBNet consta de tres partes: extracción de características, modelo de alineación de características de iluminación (IAFA) y, por último, un mecanismo de iluminación. El módulo IAFA consiste en dos capas convolucionales y tres capas totalmente conectadas y adapta el modelo a diferentes condiciones de iluminación. El módulo IAFA minimiza la función de entropía cruzada entre los valores de iluminación predichos y los reales. En (Mohanty S., et al., 2019) se detectan objetos en movimiento en ambientes con variaciones de iluminación al extraer una característica invariante a la iluminación. Las premisas que se siguen en ese trabajo es que la intensidad de un pixel es el producto de la reflectancia y la iluminación de ese pixel, y que la iluminación en una región local de la imagen es suave, es decir, el valor de la iluminación de dos pixeles cercanos es parecido. S. Kumar et. al. (Kumar S., et al., 2021) realizan una detección del plano principal de una imagen, el cual es robusto ante variaciones de la iluminación. Para esto se extraen características de intensidad, color, homogeneidad local y entropía. Una vez que se tiene el vector de características, durante el modelado del fondo, se utiliza la distancia de Canberra para medir la similitud entre la media del vector de características del cuadro actual con el modelo. Finalmente, un algoritmo de selección adaptativo selecciona un valor de umbral para clasificar un pixel como fondo o plano principal.

En (Dai X., et al., 2021). se presenta la red TIRNet para la detección de objetos en imágenes infrarrojas térmicas (TIR). La utilización de las cámaras que obtienen imágenes TIR es debido a que las imágenes IR raramente son influenciadas por los cambios en la iluminación, así que, desde el tipo de imágenes utilizada, se soluciona en gran medida la detección de objetos ante variaciones en la iluminación. La TIRNet utiliza como backbone la red VGG, y una rama residual para obtener características más robustas para la regresión y clasificación. A partir de estas dos premisas se obtiene la invarianza ante la iluminación. M. Afifi et. al (Afif M., et al., 2020) detectan objetos en interiores utilizando el detector RetinaNet con varias redes como backbone, tales como ResNet, DenseNet y VGGNet. Este procedimiento presenta una alta capacidad de detección de los objetos interiores en condiciones de iluminación desafiantes. En la Tabla se resume la información recientemente descrita.

4.5 Objetos pequeños y cambios de escala

Este problema consiste en la dificultad de detectar objetos pequeños en una imagen, así como detectar el mismo objeto si aparece en diferentes tamaños.

Artículo	Método	Artículo	Método
(Wu D., et al., 2020)	El conjunto de entrenamiento tiene muestras con distinta iluminación.	(Mohanty S., et al., 2019) y (Kumar S., et al., 2021)	Extracción de características robustas ante cambios de iluminación.
(Wu D., et al., 2020)	El conjunto de entrenamiento tiene muestras con distinta iluminación.	(Mohanty S., et al., 2019) y (Kumar S., et al., 2021)	Extracción de características robustas ante cambios de iluminación.
(Olawajun M., 2021; Badeka E., et al., 2020; Li G., et al., 2020; Wu Y., et al., 2020)	El conjunto de entrenamiento cuenta con imágenes con distinta iluminación.	(Dai X., et al., 2021)	Se utilizan imágenes térmicas.
(Xiao B., et al., 2021)	Se estima la distribución de la iluminación y se reconstruyen los detalles.	(Afif M., et al., 2020)	Se utiliza RetinaNet.
(Zhou K., et al., 2020)	Módulo de alineación de características de iluminación que minimiza una función de pérdida entre los valores de iluminación predichos y los reales.		

Tabla 7. Resumen de los artículos que resuelven el problema de la iluminación.

La idea más utilizada en las diversas metodologías es la combinación de características multiescala, que ayudan a obtener características del mismo objeto a diferentes tamaños y de información contextual (Bosquet B., et al., 2020; Leng J., et al., 2021; Zheng Q. y Chen Y., 2021; Zhang Y., et al., 2019; Zheng H., et al., 2020; Yan Z., et al., 2021; Li L., et al., 2021; Zhang S., et al., 2020; Fan M., et al., 2021; Luo H., et al., 2019; Ji S., et al., 2023). También hay metodologías que utilizan conjuntos de datos donde existen representaciones de diferente tamaño de los objetos de interés (Zhang H., et al., 2021; Kisantal M., et al., 2019). Una idea menos explorada ha sido la de predecir valores de píxeles para la detección de objetos pequeños (Lee G., et al., 2021) o el uso de Transformers (Xu S., Gu J., et al., 2023). A continuación, se describe trabajos que no solo atienden el problema de detección de objetos pequeños sino también cambios de escala.

Bosquet et. al (Bosquet B., et al., 2020) introducen la red neuronal convolucional STDnet, la cual es una red enfocada en la detección de objetos pequeños (los autores establecen un objeto pequeño como aquel que es menor a 16x16 píxeles). La red STDnet engloba 5 etapas: convoluciones iniciales, Region Context Network (RCN), convoluciones tardías, Region Proposal Network (RPN), y el clasificador. En las convoluciones iniciales la red STDnet aprende características básicas, luego, en la RCN se seleccionan las regiones con mayor probabilidad de contener objetos pequeños. Después se obtiene un mapa de características de las convoluciones iniciales, pero solamente para las regiones seleccionadas por la RCN.

A continuación, se extraen características con más información semántica en las convoluciones tardías, estas características son la entrada a la etapa de RPN, la cual localiza los objetos con mayor precisión. Finalmente, los objetos localizados son clasificados en su respectiva clase. Además de proponer la red STDnet, se presenta un conjunto de muestras, USC-GRAD-STDdb, que consiste en más de 56000 objetos pequeños etiquetados en escenarios desafiantes. Así pues, tanto con la red STDnet como con USC-GRAD-STDdb abordan el problema de la detección de objetos pequeños. La red profunda utilizada como backbone es la red ResNet-50. J. Leng et. al (Leng J., et al., 2021) proponen la red IENet (Internal-External Network), la cual utiliza tanto la apariencia del objeto pequeño como la información contextual para una detección más robusta. La IENet consiste en tres partes: aumento de características, generación de propuestas y clasificación. La etapa de aumento de características tiene un módulo, Bi-FMM (Bidirectional Feature Fusion Module), el cual detecta características internas del objeto. La etapa de generación de propuestas tiene el módulo CRM (Context Reasoning Module), el cual mejora la calidad de las generaciones de propuestas utilizando información contextual.

Por último, la etapa de clasificación incluye el módulo CFAM (Context Feature Augmentation Module), este módulo aprende relaciones entre las propuestas dadas por el CRM, y esas relaciones son utilizadas para producir información global asociada con las regiones propuestas para mejorar la clasificación. La red IENet fue evaluada con los conjuntos de datos MS COCO y WIDER FACE. Q. Zheng e Y. Chen (Zheng Q. y Chen Y., 2021) proponen una estrategia para el mejoramiento de características multiescala. El enfoque que ellos utilizan tiene dos partes: 1) el módulo multi-scale auxiliary enhancement network, el cual mejora las características al introducir características detalladas de bajo y medio nivel de la red original (VGG-16), y el módulo Adaptive Interaction, el cual mejora el desempeño del modelo al utilizar las características mejoradas. De este modo, al utilizar múltiples escalas, obtiene un buen desempeño en la detección de objetos pequeños. En (Zhang Y., et al., 2019) se enfocan en la detección de rostros que aparecen con un tamaño pequeño en una imagen. El método en ese trabajo consiste en generar una imagen clara y de alta resolución de un rostro con una red GAN, a partir de una imagen pequeña y borrosa. Además, se utiliza información contextual en combinación con la red GAN propuesta. La información contextual se utiliza debido a que los seres humanos, al buscar un rostro, no solamente se enfocan en el rostro, sino que buscan cabello, cuello, incluso el cuerpo, sobre todo en situaciones donde es difícil encontrar un rostro. H. Zheng et al (Zheng H., et al., 2020) abordan la detección de objetos a diferentes escalas de la siguiente manera: primeramente, se obtienen características basadas en gradientes orientados para obtener orientaciones locales discriminativas.

Luego, se utiliza la convolución dilatada para aumentar la resolución espacial del mapa de características, lo que ayuda a la detección multiescala. Z. Yan et al. (Yan Z., et al., 2021) proponen un detector de una etapa llamado LocalNet. LocalNet tiene como backbone la red long neck ResNet, la cual preserva información más detallada en las etapas iniciales para mejorar la representación de objetos pequeños. Esto debido a que mejora los campos receptivos de las relaciones de aspecto múltiple (multiescala) durante la extracción de características.

En (Li L., et al., 2021) se emplea la red Attention Feature Pyramid Transformer Network (AFPN), la cual aprende a detectar objetos multiescala mediante unos mapas de características multiescala. S. Zhang et. al (Zhang S., et al., 2020) presentan la red AMS-Net, Assymetry multi-stage nerowrk, para la detección de peatones a diferentes escalas. En términos generales, la AMS-Net está compuesta por tres etapas: la etapa 1 tiene como entrada una imagen piramidal, esto es, una misma imagen a múltiples escalas diferentes, a la cual se le extraen características multiescala e información contextual para obtener las regiones de propuesta (donde puede estar un objeto); en la etapa 2 todas las regiones de propuesta entran a una CNN que utiliza un mecanismo de aumento de características globales para mejorar la clasificación, además, en esta etapa se rechazan muchos falsos positivos.

Finalmente, en la etapa 3 se mejora la localización de los objetos con una red más profunda. M. Fan et. al (Fan M., et al., 2021) exponen un método para la detección de objetos pequeños en imágenes infrarrojas. El método consiste en mejorar la intensidad de los objetos con características de intensidad locales, luego se realiza una detección de esquinas. A continuación, las regiones donde posiblemente haya un objeto entran al clasificador (CNN). En (Luo H., et al., 2019) se utiliza un algoritmo que fusiona información contextual y la red YOLOV3 llamado Contextual-YOLOV3, la detección de objetos pequeños se lleva a cabo al combinar una FPN multiescala con información contextual. S. Ji et. al. (Ji S., et al., 2023) proponen un modelo, MCS-YOLO v4, para la detección de objetos pequeños. Este modelo utiliza información contextual multiescala y se propone la función de costo Soft-CIOU. Las características de los objetos pequeños se combinan con características contextuales para hacer la extracción de características más robusta. La función de costo Soft-CIOU se utiliza porque en el proceso de ajustar los rectángulos delimitadores pequeños cambios tienen un gran impacto (ya que los objetos a detectar son pequeños).

H. Zhang et. al (Zhang H., et al., 2021) generan una base de datos en la cual se introducen numerosas muestras de imágenes UAV a diferentes distancias para que los objetos aparezcan con un tamaño distinto. De esta manera, algún detector podría detectar, al ser entrenado con esa base de datos, la misma clase de objeto a diferentes escalas. Similarmente, M. Kisantal et. al (Kisantal M., et al., 2019) realizan un aumento de datos en el conjunto MS COCO para que haya más objetos pequeños en las imágenes.

G. Lee et. al (Lee G., et al., 2021) comparten la red FEN (feature enhancement network) basada en la arquitectura U-Net, con el objetivo de enriquecer las características de las regiones donde están los objetos pequeños en imágenes que tienen ruido. Primeramente, se extraen características de imágenes con ruido, luego, de manera aleatoria, se eliminan algunos valores de esas las características extraídas. Y son estas características la entrada de la FEN. El objetivo de la FEN es predecir los valores eliminados de las características utilizando los valores de los pixeles cercanos. Es esto lo que permite obtener mejores características en objetos pequeños presentes en imágenes con ruido.

En (Xu S., Gu J., et al., 2023) se propone la red DKTNNet. Esta red utiliza las redes ResNet50 y FPN para extraer características de la imagen de entrada. Estas características son la entrada de un dual-key transformer para aumentar la correlación entre Q, query, y V, value, lo cual mejora la discriminación de las características. En la Tabla 7 se presenta un resumen de la subsección sobre detección de objetos pequeños.

Artículo	Método	Artículo	Método
(Bosquet B., et al., 2020)	Se presenta la red enfocada en la detección de objetos pequeños Red STDnet, así como el conjunto de datos USC-GRAD-STDdb, con 56000 objetos pequeños etiquetados en ambientes desafiantes.	(Zhang S., et al., 2020; Luo H., et al., 2019)	Información contextual junto con características de múltiples escalas.
(Leng J., et al., 2021)	Red <i>Internal-External Network</i> (IENet), que utiliza tanto las características del objeto como información contextual.	(Fan M., et al., 2021)	Características de intensidad local y detección de esquinas en imágenes infrarrojas.
(Zheng Q. y Chen Y., 2021)	Características de múltiples escalas.	(Ji S., et al., 2023)	Información contextual y la función de costo Soft-CIOU.
(Zhang Y., et al., 2019)	Combinación de una red GAN (que obtiene una imagen de alta resolución) con información contextual para la detección de rostros.	(Zhang H., et al., 2021)	Base de datos de imágenes UAV a diferentes alturas.
(Zheng H., et al., 2020)	Convolución dilatada para aumentar la resolución espacial del mapa de características.	(Kisantal M., et al., 2019)	Aumento de datos en el conjunto MS COCO para que aparezcan más objetos pequeños.
(Yan Z., et al., 2021)	Mejoramiento de los campos receptivos de las relaciones de multiescala.	(Lee G., et al., 2021)	Red <i>feature enhancement newtork</i> (FEN), la cual mejora las características de las regiones donde se ubican los objetos pequeños.
(Li L., et al., 2021)	Características multiescala.	(Xu S., Gu J., et al., 2023)	<i>Transformer</i> para mejorar la discriminación de las características.

Tabla 7. Resumen de los artículos que atienden el problema de detección de objetos pequeños.

4.6 Variación entre clases y la misma clase

Este problema consiste en que puede haber dos clases distintas que se asemejen, o bien, una misma clase con objetos que se aprecien muy diferentes. Diversos estudios se han realizado para resolver este problema, donde la metodología más utilizada ha sido la modificación de las funciones de pérdida de las redes profundas (Sun B., et al., 2021; Yuan J., et al., 2020; Li C., et al., 2019) para maximizar las diferencias entre las clases. El uso de nuevos paradigmas ha sido de gran utilidad, como Few-Shot (Cao Y., et al., 2021; Li B., et al., 2021), en el cual puede realizar una correcta detección aun con pocas muestras. Una discriminación que parte de lo general a lo específico ha sido estudiada en (Shin D., et al., 2020), la idea central son los árboles de decisión. El aumento del conjunto de datos también ha sido de utilidad para resolver este problema (Li K., et al., 2020). Algunos de los trabajos relacionados se comentan a continuación.

En (Sun B., et al., 2021) se propone una función de pérdida, contrastive proposal encoding loss, la cual mide la similitud semántica entre las regiones de propuestas. X^+ representa las propuestas positivas, es decir, las regiones de la misma categoría (mismo objeto) y X^* corresponde a las regiones negativas, que son aquellas que tienen diferentes categorías. La función de pérdida debe ser tal que $\text{score}(f(X), f(X^+)) > \text{score}(f(X), f(X^*))$, es decir, se debe tener una varianza pequeña entre la misma categoría y mayor entre clases. Algo parecido se presenta en (Yuan J., et al., 2020) donde se utiliza una CNN para la detección de objetos. En ese trabajo se presenta una función de pérdida llamada Inter-class loss, la cual ayuda al algoritmo a aprender información discrepante entre categorías, lo cual hace que detecte de mejor manera objetos de la misma categoría. C. Li et. al. (Li C., et al., 2019), modifican la función de costo de la red YOLOv3, de una función logística a una softmax para maximizar la diferencia de las características entre las clases.

Y. Cao et. al. (Cao Y., et al., 2021) y B. Li et. al. (Li B., et al., 2021) presentan trabajos en los cuales se detectan objetos de clases nuevas a partir de pocas muestras mediante el enfoque de aprendizaje Few-Shot, y así resuelven el problema de la varianza entre clases. Y. Cao et. al. (Cao Y., et al., 2021) resuelven el problema de la varianza entre clases mediante asociación y discriminación. Durante la asociación se construye un espacio de características que imita un espacio de características de una clase base, el nuevo espacio se asocia a la clase base de acuerdo con su similitud semántica. En la discriminación se asegura la separabilidad de las clases con un margen de pérdida. En B. Li et. al. (Li B., et al., 2021) se presenta un margen de equilibrio de clases (CME) que optimiza tanto la partición del espacio de características y la reconstrucción de las clases nuevas. CME reserva un espacio adecuado de las nuevas clases al utilizar un margen de pérdida durante el aprendizaje.

En (Shin D., et al., 2020) se utiliza un enfoque basado en árboles de decisión donde la primera rama detecta clases amplias, esto quiere decir clases como «animales» o «vehículos», de esta manera se separa la diferencia entre clases, luego se detecta la clase de los objetos para cada nodo (clase), esto es, por ejemplo, de la clase «animales» surgen ramas distintas para la clase «perro», «gato», etc. Finalmente, se tienen subclases donde se detecta un único objeto de la clase.

K. Li et. al (Li K., et al., 2020) presentan en su artículo un conjunto de muestras de entrenamiento que, entre otras cualidades, tiene una adición de muestras para aumentar la diversidad entre la misma clase. En la Tabla 8 se recapitula la información anteriormente escrita.

Artículo	Método	Artículo	Método
(Sun B., et al., 2021; Yuan J., et al., 2020; Li C., et al., 2019)	Se propone una función de pérdida.	(Shin D., et al., 2020)	Enfoque basado en árboles de decisión.
(Cao Y., et al., 2021; Li B., et al., 2021)	Construcción de un mapa de características.	(Li K., et al., 2020)	Mejoramiento del conjunto de entrenamiento.

Tabla 8. Resumen de los artículos que atienden el problema de detección de objetos con variación entre la misma clase y entre diferentes clases.

4.7 Deformación y cambios de pose

Este problema consiste en que, si un objeto cambió de posición o se ve deformado por cualquier razón, podría provocar que un sistema de visión ya no sea capaz de detectarlo de manera correcta.

La tendencia para resolver este problema está en el aprender múltiples perspectivas del objeto (Zhang X., et al., 2022; Fang F., et al., 2019) y la extracción de características en regiones que correspondan a la forma real del objeto (Mordan T., et al., 2019; Zhang P., et al., 2020; Liu Y., et al., 2021; Dai J., et al., 2017). Es decir, las ideas de estas dos metodologías son similares: extraer características que ayuden a tener un mayor conocimiento de las formas de los objetos. Estos enfoques resultan de mayor interés para resolver este problema que otros presentados en la literatura (Das A., et al., 2022). A continuación, se presentan los trabajos consultados que resuelven el problema de deformación y cambios de pose.

X. Zhang et al. (Zhang X., et al., 2022) presentan una red sensible a la perspectiva, PSNet, para la detección de objetos. La red PSNet aprende múltiples espacios de perspectiva, donde en cada uno de estos, las características semánticas son desacopladas para hacerla robusta ante cambios de perspectiva.

El cambio de perspectiva indica que el objeto aparece en diversas poses, rotaciones, etcétera; con lo cual, es una manera de solucionar el problema de detectar un objeto que se vea diferente en distintas imágenes. En (Fang F., et al., 2019) se presenta un enfoque híbrido que combinan la red Faster R-CNN con un modelo de agrupamiento para la detección de objetos alargados. T. Mordan et. al (Mordan T., et al., 2019) parten de la base de que normalmente, en la detección de objetos, se utilizan rectángulos delimitadores, de los cuales se extraen características sin importar la forma real del objeto, debido a esto introducen la red DP-FCN, la cual aplica deformaciones en las regiones para aprender mejores representaciones que se ajusten de una manera más adecuada a la forma real del objeto.

Esta red tiene dos contribuciones importantes: invarianza a transformaciones locales y la obtención de información geométrica para describir más fielmente a los objetos. P. Zhang et. al (Zhang P., et al., 2020) detectan y siguen objetos sobresalientes, su enfoque utiliza mapas sobresalientes espaciotemporales, y para lidiar con objetos sensibles a escalas (y, por extensión de acuerdo con los autores, a objetos no rígidos) se implementa una FCN. Enfocándonos en la parte del contexto deformable, se utiliza una convolución deformable para la localización de los rectángulos delimitadores de objetos no rígidos presentada en (Dai J., et al., 2017).

Esta convolución es embebida en la estructura multiescala añadiendo un desplazamiento 2D de la convolución estándar, lo cual resulta en estructuras adaptativas a las formas geométricas en los kernels de convolución. Y. Liu et. al (Liu Y., Duanmu M., et al., 2021) proponen el módulo Multi-Scale Deformation Module (MSDM) para obtener pistas visuales multiescala y de formas variantes de objetos sobresalientes. Los autores utilizan una modificación de la red ResNet-50 como backbone. MSDM consiste en dos pasos: el primero es la extracción de contexto multiescala, donde se aprenden características multiescala, y el segundo es la extracción del contexto deformable. Combinando ambos pasos, el MSDM es capaz de capturar información a diferentes escalas y diversas formas.

En (Das A., et al., 2022) se presenta un detector de peatones en diferentes poses en donde se extraen características de peatones y se realiza una segmentación por instancias, después un codificador aprende características específicas en peatones dependiendo de su pose. En la Tabla se resume la información presentada sobre la detección de objetos ante deformaciones y cambios de pose.

5. Conclusiones y trabajo futuro

La detección de objetos es una de las etapas más importantes de visión por computadora. Por esta razón, se han desarrollado una gran cantidad de modelos de detección de objetos que presentan buenos resultados en ciertos conjuntos de muestras, en clases específicas, en ambientes restringidos, y utilizando

Artículo	Método	Artículo	Método
(Zhang X., et al., 2022)	Red PSNet, que aprende diferentes perspectivas.	(Liu Y., Duanmu M., et al., 2021)	Convolución deformable para la localización de objetos de no rígidos.
(Fang F., et al., 2019)	Red DP-FCN, que aplica transformaciones en las regiones en vez de utilizar solamente rectángulos delimitadores.	(Fang F., et al., 2019)	Combinación de la red Faster R-CNN con un módulo de agrupamiento.
(Zhang P., et al., 2020)	Mapas de características sensible a múltiples escalas.	(Das A., et al., 2022)	Segmentación por estancias y extracción características de peatones en función de su pose.

Tabla 10. Resumen de los artículos que atienden el problema de detección de objetos ante deformaciones y cambios de pose.

e implementado una de entre tantas ideas y posibilidades. Sin embargo, la detección de objetos sigue siendo una tarea compleja, ya que presenta diversos retos como los mencionados en este artículo.

A partir del análisis realizado en este trabajo, resulta evidente que se debería incrementar el número de clases a detectar, pues existen aplicaciones en las cuales se tienen varias clases de interés. También se observó que, para resolver algunos retos como la oclusión, cambios de iluminación o detección de objetos pequeños y cambios de escala es típico aumentar el conjunto de datos para hacer al detector más robusto; sin embargo, sería más interesante y contribuiría más el aportar y aplicar ideas en los modelos que solucionen dichos problemas.

Por otra parte, como resultado del análisis de la detección de objetos con confusión, resulta evidente que es fundamental la detección en ambientes más diversos y distintos, y también es crucial el estudio de las relaciones del fondo con los objetos de interés. Además, debido a que la tendencia es la detección de bordes, el desarrollo en esta área será de gran utilidad.

Basándonos en el análisis de la detección de objetos utilizando información contextual se desprende que, a pesar de su correcta aplicación, es indispensable asegurar que exista una relación útil entre objetos y/o el objeto y el fondo, ya que no suele presentarse este análisis en los trabajos reportados en este artículo y es la relación que existe entre elementos de la imagen lo que permite utilizar la información contextual. Esta relación no solamente debe darse entre los objetos, sino en atributos presentes en la imagen (por ejemplo, si hubiera una relación entre el origen de una fuente de iluminación y la manera en que un objeto proyecta una sombra, se debería analizar si esta relación sirve como información contextual).

Considerando los métodos para detectar objetos con variación entre clases y entre la misma clase, se concluye que sería benéfico el uso de otros paradigmas de aprendizaje para la generalización entre clases y subclases, que, a pesar de no ser la metodología más seguida, es la que mayor impacto podría tener.

De acuerdo con las ideas presentadas en la detección de objetos con deformaciones y cambios de pose se concluye que la forma en la que el modelo extrae características que se asemejen más a la forma del objeto es un área interesante de investigación. Debido a esto, la implementación de redes profundas que encuentren relaciones espaciales (como las redes tipo cápsula) sería un gran beneficio para resolver este problema.

Con base en el análisis presentado en este artículo se mantiene el énfasis en que existen varias soluciones a un mismo problema, y como se aprecia en diversos artículos publicados, algunos solucionan un problema al resolver otro. Con lo cual se hace evidente que la integración es un proceso fundamental en la detección de objetos.

Por lo anterior, concluimos que un posible tema de la investigación futura en la detección de objetos se debería concentrar, entre otros enfoques y/o métodos que se comentan en diversos estudios (como los mencionados en la sección 2), en tratar de adoptar ideas y mecanismos realizados por el ser humano al detectar objetos. Esto se propone porque la visión humana tiene un buen desempeño para reconocer patrones, obtener información del contexto y localizar objetos en escenas complicadas.

Agradecimientos

Los autores agradecen el financiamiento del Tecnológico Nacional de México/I.T. Chihuahua bajo el proyecto 16431.23-P para realizar este proyecto.

REFERENCIAS

- Abbas S., et al. (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, 103514.
- Adouani A., et al. (2019). Comparision of Haar-like, HOG and LBP approaches for face detection in video sequences. *2019 16th International Multi-Conference on Systems, Signals & Devices (SSD)* (págs. 266-271). Estambul, Turquía: IEEE.
- Adreopoulos A. y Tsotsos J. (2013). A Computational Learning Theory of Active Object Recognition Under Uncertainty. *International Journal of Computer Vision*, 95-143.
- Afif M., et al. (2020). An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation. *Neural Processing Letters*, 2265-2279.
- Ajmera F., et al. (2021). Survey on Object Detection in Aerial Imagery. *IEEE Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (págs. 1050-1055). Tirunelveli, India: IEEE.
- Ansari S. (2019). A Review on SIFT and SURF for Underwater Image Feature Detection and Matching. *2019 IEEE International Conference on Electrical, Computer and Communication Technologies* (págs. 1-4). IEEE.
- Arfi A., et al. (2020). Real Time Human Face Detection and Recognition Based on Haar Features. *2020 IEEE Region 10 Symposium (TENSYP)* (págs. 517-521). IEEE.
- Arnold E., et al. (2019). A Survey on 3D Object Detection Methods for Autonomous Driving Applications. *IEEE Transactions on Intelligent Transportation Systems*, 3782-3795.
- Arulprakash E. y Aruldoss M. (2021). A study on generic object detection with emphasis on future research. *Journal of King Saud University-Computer and Information Sciences*.
- Arunmozhi A. y Park J. (2018). Comparision of HOG, LBP and Haar-Like Features for on-Road Vehicle Detection. *2018 IEEE International Conference on Electro/Information Technology (EIT)* (págs. 362-367). IEEE.
- Aziz L., et al. (2020). Exploring Deep Learning-Based Architecture, Strategies, Applications and Current Trends in Generic Object Detection: A Comprehensive Review. *IEEE Access*, 170461-170495.
- Badeka E., et al. (2020). Harvest Crate Detection for Grapes Harvesting Robot Based on YOLOv3 Model. *2020 Fourth International Conference on Intelligent Computing in Data Sciences (ICDS)*, (págs. 1-5).
- Bastanfard A., et al. (2019). Improving Tracking Soccer Players in Shaded Playfield Video. *2019 5th Iranina Conference on Signal Processing and Intelligent Systems (ICSPIS)* (págs. 1-8). IEEE.
- Bay H., et al. (2006). SURF: Speeded Up Robust Features. *European Conference on Computer Vision (ECCV)* (págs. 404-417). Springer.
- Biswas R., et al. (2021). A new perceptual hashing method for verification and identity classification of occluded faces. *Image and Vision Computing*, 104186.
- Biswas S., et al. (2021). Beyond document object detection: instance-level segmentation of complex layouts. *International Journal on Document Analysis and Recognition*, 269-281.
- Bochkovskiy A., et al. (2020). YoloV4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.
- Borji A., et al. (2019). Salient object detection: A survey. *Computational Visual Media*, 117-150.
- Bosquet B., et al. (2020). STDnet: Exploiting high resolution feature maps for small object detection. *Engineering Applications of Artificial Intelligence*, 103615.

- Cao Y., et al. (2021). Few-Shot Object Detection Via Association and Discrimination. *Advances in Neural Information Processing*.
- Cen F., et al. (2021). Deep feature augmentation for occluded image classification. *Pattern Recognition*, 107737.
- Chapel M. y Bouwmans T. (2018). New trends on moving object detection in video images captured by a moving camera: A survey. *Computer Science Review*, 157-177.
- Chapel M. y Bouwmans T. (2020). Moving objects detection with a moving camera: A comprehensive review. *Computer Science Review*, 100310.
- Chen G., et al. (2022). Camouflaged Object Detection via Context-aware Cross-level Fusion. *IEEE Transactions on Circuit Systems and Systems for Video Technology*.
- Chen T., et al. (2022). Boundary-guided network for camouflaged object detection. *Knowledge-Based Systems*, 108901.
- Chen Z., et al. (2020). Underwater salient object detection by combining 2D and 3D visual. *Neurocomputing*, 239-259.
- Cheng G., et al. (2022). Towards Large-Scale Small Object Detection: Survey and Benchmarks. *arXiv preprint arXiv:2207.13096*.
- Chilukuri D., et al. (2022). A robust object detection system with occlusion handling for mobile devices. *Computational Intelligence*.
- Cyger S. y Czyzewski A. (2020). Towards Robust Pedestrian Detection With Data Augmentation. *IEEE Access*, 126674-126683.
- Dai J., et al. (2016). R-fcn: Object detection via region-based fully convolutional networks. *Advances in neural information processing systems*, 379-387.
- Dai J., et al. (2017). Deformable convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision*, (págs. 764-773).
- Dai K., et al. (2019). Visual Tracking via adaptive spatially-regularized correlation filters. *Proceedings of the IEEE conference on computer vision and pattern recognition* (págs. 4670-4679). IEEE.
- Dai X., et al. (2021). TIRNet: Object detection in thermal infrared images for autonomous driving. *Applied intelligence*, 1244-1261.
- Dalal N. y Triggs B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (págs. 886-603). IEEE.
- Das A., et al. (2022). Deep Multi-Task Networks For Occluded Pedestrian Pose Estimation. *arXiv preprint arXiv:2206.07510*.
- Dash P. y Sigappi A. (2018). Detection and Classification of Retinal Diseases in Spectral Domain Optical Coherence Tomography Images based on SURF descriptors. *2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA)* (págs. 1-6). IEEE.
- Deng J., et al. (2020). A review on research on object detection based on deep learning. *Journal of Physics: Conference Series*, 12028.
- Diwan T., et al. (2023). Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications*, 9243-9275.
- Dong J., et al. (2020). Occlusion-Aware GAN for Face De-Occlusion in the Wild. *2020 IEEE International Conference on Multimedia and Expo (ICME)* (págs. 1-6). IEEE.

- Dong W., et al. (2021). Ellipse R-CNN: Learning to Infer Elliptical Object from Clustering and Occlusion. *IEEE Transactions on Image Processing*, 2193-2206.
- Du S. y Wang S. (2022). An Overview of Correlation-Filter-Based Object Tracking. *IEEE Transactions on Computational Social Systems*, 18-31.
- Espinosa J., et al. (2020). Detection of Motorcycles in Urban Traffic Using Video Analysis: A Review. *IEEE Transactions on Intelligent Transportation Systems*, 6115-6130.
- Fan M., et al. (2021). Infrared small target detection based on region proposal and CNN classifier. *Signal, Image and Video Processing*, 1927-1936.
- Fang F., et al. (2019). Combining Faster R-CNN and Model-Driven Clustering for Elongated Object Detection. *IEEE Transactions on Image Processing*, 2052-2065.
- Feng Y., et al. (2021). Detect Faces Efficiently: A Survey and Evaluations. *Transactions on Biometrics, Behavior, and Identity Source*.
- Flores M., et al. (2019). Pedestrian Detection Under Partial Occlusion by using Logic Inference, HOG and SVM. *IEEE Latin America Transactions*, 1552-1559.
- Geetha S., et al. (2021). Machine Vision Based Fire Detection Techniques: A Survey. *Fire Technology*, 591-623.
- Girshick R. (2015). Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (págs. 1440-1448). IEEE.
- Girshick R., et al. (2014). Rich feature hierarchies for accurate object detection and semantics segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition* (págs. 580-587). IEEE.
- Gkioxioari G., et al. (2019). Mesh r-cnn. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (págs. 9785-9795). IEEE.
- Gong Y., et al. (2020). Context-Aware Convolutional Neural Network for Object Detection in VHR Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 34-44.
- Guo J., et al. (2020). Object detector with enriched global context information. *Multimedia Tools and Applications*, 29551-29571.
- Guo T., et al. (2020). Detection of Ice Hockey Players and Teams via a Two-Phase Cascaded CNN Model. *IEEE Access*, 195062-195073.
- Gupta S., et al. (s.f.). Improved object recognition results using SIFT and ORB feature detector. *Multimedia*.
- Gurbina M., et al. (2019). Tumor Detection and Classification of MRI Brain Image using Different Wavelet Transforms and Support Vector Machines. *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)* (págs. 505-508). IEEE.
- Hajizadeh M., et al. (2023). MobileDenseNet: A new approach to object detection on mobile devices. *Expert Systems with Applications*.
- Han B., et al. (2020). Small-Scale Pedestrian Detection Based on Deep Neural Network. *IEEE Transactions on Intelligent Transportation Systems*, 3046-3055.
- Han J., et al. (2018). Advanced Deep-Learning Techniques for Salient and Category-Specific Object Detection: A Survey. *IEEE Signal Processing Magazine*.
- Han L., et al. (2021). Context and Structure Mining Network for Video Object Detection. *International Journal of Computer Vision*, 2927-2946.

- He K., et al. (2015). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1904-1916.
- He K., et al. (2017). Mask r-cnn. *Proceedings of the IEEE international conference on computer vision* (págs. 2961-2969). IEEE.
- Heo J., et al. (2022). Occlusion-aware spatial attention transformer for occluded object recognition. *Pattern Recognition Letters*, 70-76.
- Huang G., et al. (2019). Ship detection based on squeeze excitation skip-connection path networks for optical remote sensing images. *Neurocomputing*, 215-223.
- Huang G., et al. (2023). A Survey of Self-Supervised and Few-Shot Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4071-4089.
- Huang X., et al. (2021). A bioinspired retinal neural network for accurately extracting small-target motion information in clutter backgrounds. *Image and Vision Computing*, 104266.
- Jarunakarint V., et al. (2020). Survey and Experimental Comparison of Machine Learning Models for Motorcycle Detection. *2020 5th International Conference on Information Technology (InCIT)* (págs. 320-325). IEEE.
- Ji G., et al. (2022). Fast Camouflaged Object Detection via Edge-based Reversible Re-calibration Network. 108414.
- Ji S., et al. (2023). An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information. *Computers and Electrical Engineering*, 108490.
- Jiao L., et al. (2019). A Survey of Deep Learning-Based Object Detection. *IEEE Access*, 128837-128868.
- Jocher G. (22 de 4 de 2020). *YOLOv5 by Ultralytics*. Obtenido de <https://github.com/ultralytics/yolov5>
- Jocher G., et al. (22 de 4 de 2023). *YOLO by Ultralytics*. Obtenido de Available: <https://github.com/ultralytics/ultralytics>
- Kaur M. y Min C. (2018). Automatic Crop Furrow Detection for Precision Agriculture. *2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS)* (págs. 520-523). IEEE.
- Kaur P., et al. (2020). A survey on brain tumor detection techniques for MR images. *Multimedia Tools and Applications*, 21771-21814.
- Kaushal M., et al. (2018). Soft computing based object detection and tracking approaches: State-of-the-Art survey. *Applied Soft Computing*, 423-464.
- Kellenberger B., et al. (2019). Half a Percent of Labels is Enough: Efficient Animal Detection in UAV Imagery Using Deep CNNs and Active Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 9524-9533.
- Kisantal M., et al. (2019). Augmentation for small object detection. *arXiv preprint*.
- Köhler M., et al. (2023). Few-Shot Object Detection: A Comprehensive Survey. *IEEE Transactions on Neural Networks and Learning Systems*, 1-21.
- Kreim A., et al. (s.f.). Using synthetic data for person tracking under adverse weather conditions. *Image and Vision Computing*, 104187.
- Kumar N. (2018). Thresholding in salient object detection: a survey. *Multimedia Tools and Multimedia*, 19139-19170.
- Kumar S., et al. (2021). An improved scheme for multifeature-based foreground detection using challenging conditions. *Digital Signal Processing*, 103030.
- Lang W., et al. (2021). A survey of 3D object detection. *Multimedia Tools and Applications*, 29617-29641.

- Lee G., et al. (2021). Self-Supervised Feature Enhancement Networks for Small Object Detection in Noisy Images. *IEEE Signal Processing Letters*, 1026-1030.
- Leng J., et al. (2021). Realize your surroundings: Exploiting context information for small object detection. *Neurocomputing*, 287-299.
- Li B., et al. (2021). Beyond Max-Margin Class Margin Equilibrium for Few-Shot Object Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (págs. 7363-7372).
- Li C. et. al. (2022). YOLOV6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv preprint arXiv:2209.02976*.
- Li C., et al. (2019). Face Detection Based on YOLOv3. *Recent Trends in Intelligent Computing, Communication and Devices*, 277-284.
- Li F., et al. (2022). Faster R-CNN with Generative Adversarial Occlusion Network for Object Detection. *2022 14th International Conference on Machine Learning and Computing (ICMLC)*, (págs. 536-541).
- Li G., et al. (2020). Faster R-CNN Deep learning Model for Pedestrian Detection from Drone Images. *SN Computer Science*.
- Li J., et al. (2019). Multi-scale HOG feature used in object detection. *Tenth International Conference on Graphics and Image Processing (ICGIP)*.
- Li K., et al. (2020). Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 296-307.
- Li L., et al. (2021). Scale-Insensitive Object Detection via Attention Feature Pyramid Transformer Network. *Neural Processing Letters*.
- Li N., et al. (2020). Detection of Animal Behind Cages Using Convolutional Neural Network. *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (págs. 242-245). IEEE.
- Li Z., et al. (2017). Light-head r-cnn: In defense of two-stage object detector. *arXiv preprint arXiv:1711.07264*.
- Li Z., et al. (2021). A survey of 3D object detection algorithms for intelligent vehicles development. *Artificial Life and Robotics*.
- Liang D., et al. (2020). Lane Detection: A Survey with New Results. *Journal of Computer Science and Technology*, 493-505.
- Liang X., et al. (2020). Small Object Detection in Unmanned Aerial Vehicle Images Using Feature Fusion and Scaling-Based Single Shot Detector With Spatial Context Analysis. *IEEE Transactions on Circuits and Systems for Video Technology*, 1758-1770.
- Lin T., et al. (2017). Feature Pyramid Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (págs. 2117-2125). IEEE.
- Lin T., Goyal P., et al. (2017). Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision* (págs. 2980-2988). IEEE.
- Liu C., et al. (2022). Double Mask R-CNN for Pedestrian Detection in a Crowd. *Mobile Information Systems*.
- Liu L., et al. (2019). Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision*, 261-3128.
- Liu S., Liu D., et al. (2021). Effective template update mechanism in visual tracking with background clutter. *Neurocomputing*, 616-625.

- Liu S., Liu D., Seivastava G., et al. (2021). Overview of correlation filter based algorithms in object tracking. *Complex and Intelligent Systems*, 1895-1917.
- Liu T., et al. (2020). Coupled Network for Robust Pedestrian Detection With Gated Multi-Layer Feature Extraction and Deformable Occlusion Handling. *IEEE Transactions on Image Processing*, 754-766.
- Liu Y., Duanmu M., et al. (2021). Exploring multi-scale deformable context and channel-wise attention for salient object detection. *Neurocomputing*, 92-103.
- Liu Y., Peng J., et al. (2021). TSingNet: Scale-aware and context-rich feature learning for traffic sign detection and recognition in the wild. *Neurocomputing*, 10-22.
- Liu Y., Sun P., et al. (2021). A survey and performance evaluation of deep learning methods for small object detection. *Expert Systems with Applications*, 114602.
- Lowe D. (1999). Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision* (págs. 1150-1157). IEEE.
- Lu W., et al. (2016). SSD: Single shot Multibox Detector. *European Conference on Computer Vision (ECCV)*, (págs. 21-37).
- Luo H., et al. (2019). Contextual-YOLOv3: Implement Better Small Object Detection Based Deep Learning. *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDI)*, (págs. 134-142).
- Luo H., et al. (2021). Object Detection Method Based on Shallow Feature Fusion and Semantic Information Enhancement. 21839-21851.
- Meda K., et al. (2021). Artificial intelligence research within reach: an object detection model to identify rickets on pediatric wrist radiographs. *Pediatric Radiology*, 782-791.
- Minaee S., et al. (2021). Going Deeper Into Face Detection: A Survey. *arXiv preprint*.
- Mittal P., et al. (2020). Deep learning-based object detection in low-altitude UAV datasets: A survey. *Image and Vision Computing*, 204046.
- Mo Y., et al. (2019). Highlight-assisted nighttime vehicle detection using a multi-level fusion network and label hierarchy. *Neurocomputing*, 13-23.
- Mohammed S., et al. (2021). A Review on Object Detection Algorithms for Ship Detection. *IEEE 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* (págs. 1-5). IEEE.
- Mohanty S., et al. (2019). A New Approach for Moving Object Detection under Varying Illumination Environments. *2019 International Conference on Information Technology (ICIT)*, (págs. 420-424).
- Moncef S. y Othman A. (2021). Efficient Techniques For Human Face Occlusions Detection and Extraction. *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*, (págs. 1-5).
- Mondal A. (2021). Camouflage design, assessment and breaking techniques: a survey. *Multimedia Systems*.
- Moniruzzaman M., et al. (2017). Deep Learning on Underwater Marine Object Detection: A survey. *Advanced Concepts for Intelligent Vision Systems*, 150-160.
- Mordan T., et al. (2019). End-to-end Learning of Latent Deformable Part-Based Representations for Object Detection. *International Journal of Computer Vision*, 1659-1679.
- Muhammad K., et al. (2018). Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing*, 30-42.

- Naji S., et al. (2019). A survey on skin detection in colored images. *Artificial Intelligence Review*, 1041-1087.
- Nguyen N., et al. (2019). A Novel Hardware Architecture for Human Detection using HOG-SVM Co-Optimization. *2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)* (págs. 33-36). IEEE.
- Ning C., et al. (2021). Survey of pedestrian detection with occlusions. *Complex & Intelligent Systems*, 577-587.
- Nyein M. y Tint T. (2021). A Review on Advanced Detection Methods in Vehicle Traffic Scenes. *IEEE 6th International Conference on Inventive Computation Technologies (ICICT)* (págs. 642-649). IEEE.
- Ojala T., et al. (1994). Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. *Proceedings of the 12th IAPR International Conference on Pattern Recognition* (págs. 582-585). IEEE.
- Olarewaju M. (2021). Tomato detection based on modified YOLOv3 framework.
- Pandya M., et al. (2020). Analysis of Deep Learning Architectures for Object Detection - A critical Review. *IEEE-HYDICON* (págs. 1-6). IEEE.
- Pathak A., et al. (2018). Deep Learning Approaches for Detecting Objects from Images: A Review. *Progress in Computing, Analytics and Networking, Advances in Intelligent Systems and Computing*.
- Pathare S., et al. (2020). Detection of Fractures in Long Bones for Trauma Centre Patients using Hough Transform. *2020 International Conference on Communication and Signal Processing (ICCSPP)* (págs. 88-91). IEEE.
- R., M. (1986). *Estados Unidos Patente nº 4567610*.
- Radhika S., et al. (2018). Determination of Degree of Damage on Building Roofs Due to Wind Disaster from Close Range Remote Sensing Images Using Texture Wavelet Analysis. *IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium* (págs. 3366-3369). IEEE.
- Rahmatulloh A., et al. (2021). Face Mask Detection Using Haar Cascade Classifier Algorithm based on Internet of Things with Telegram Bot Notification. *2021 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, (págs. 1-5).
- Redmon J. y Farhadr A. (2017). YOLO9000: Better, Faster, Stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pág. 72637271). IEEE.
- Redmon J. y Farhadr A. (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv: 1804.02767*.
- Redmon J., et al. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (págs. 779-788). IEEE.
- Ren S., et al. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1137-1149.
- Rodríguez A., et al. (2020). Reconocimiento de la denominación de billetes a través de una aplicación móvil con reconocimiento de imagen. *ReCIBE. Reviste electrónica de Computación, Informática, Biomédica y Electrónica*, 1-16.
- Rublee E., et al. (2011). ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision (ICCV)* (págs. 2564-2571). IEEE.
- Saez A., et al. (2019). Statistical Detection of Colors in Dermoscopic Images With a Texton-Based Estimation of Probabilities. *Journal of Biomedical and Health Informatics*, 560-569.
- Said N., et al. (2019). Natural disasters detection in social media and satellite imagery: a survey. *Multimedia Tools and Applications*, 31267-21302.

- Seemanthini K. y Manjunath S. (2018). Human Detection and Tracking using HOG for Action Recognition. *Procedia Computer Science*, 1317-1326.
- Sermanet P., et al. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv: 1312.6229*.
- She X., et al. (2021). ScieNet: Deep learning with spike-assisted contextual information extraction. *Pattern Recognition*, 108002.
- Shetty A., et al. (2021). A Review: Object Detection Models. *IEEE 6th International Conference for Convergence in Technology (I2CT)* (págs. 1-8). IEEE.
- Shi P., et al. (2023). Object Detection Based on Swin Deformable Transformer-BiPAFPN-YOLOX. *Computational Intelligence and Neuroscience*.
- Shih H. y Chen J. (2021). Multiskin Color Segmentation Through Morphological Model Refinement. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 225-235.
- Shin D., et al. (2020). Dynamic MLML-tree based adaptive object detection using heterogeneous data distribution. *Multimedia Tools and Applications*, 6689-6708.
- Shou X., et al. (2019). Automated Visual Inspection of Glass Bottle Bottom With Saliency Detection and Template Matching. *IEEE Transactions on Instrumentation and Measurements*, 4253-4267.
- Singh A., et al. (2020). Animal Detection in Man-made Enviroments. *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)* (págs. 1427-1438). IEEE.
- Siris A., et al. (2021). Scene Context-Aware Salient Object Detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (págs. 4156-4166). IEEE.
- Skoviera R., et al. (2018). Object recognition in clutter color images using Hierarchical Temporal Memory combined with salient-region detection. *Neurocomputing*, 172-183.
- Song C., et al. (2020). Automatic Detection and Image Recognition of Precision Agriculture for Citrus Diseases. *2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)* (págs. 187-190). IEEE.
- Song X., et al. (2020). Progressive Refinement Network for Occluded Pedestrian Detection. *European Conference on Computer Vision (ECCV)*, (págs. 32-48).
- Song X., et al. (2022). PRNet++: Learning towards generalized occluded pedestrian detection via progressive refinement network. *Neurocomputing*, 98-115.
- Sun B., et al. (2021). FSCE: Few-Shot Object Detection via Constrastive Proposal Encoding. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (págs. 7352-7362). IEEE.
- Sun N., et al. (2006). Gender Classification Based on Boosting Local Binary Pattern. *International Symposium on Neural Networks*, (págs. 194-201).
- Suruliandi A., et al. (2012). Local binary pattern and its derivatives for face recognition. *IET Computer Vision*, 480-488.
- Terven J. y Cordova D. (2023). A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond. *arXiv preprint arXiv:2304.004501*.
- Tong K., et al. (2020). Recent advances in small object detection based on deep learning: A review. *Image and Vision Computing*, 103910.
- Ullah I., et al. (2020). A brief survey of visual saliency detection. *Multimedia Tools and Applications*, 34605-34645.

- Vashisht M. y Kumar B. (2020). A Survey Paper on Object Detection Methods in Image Processing. *International Conference on Computer Science, Engineering and Applications (ICCSEA)* (págs. 1-4). IEEE.
- Viola P. y Jones M. (2004). Robust Real-Time Face Detection. *International Journal of Computer Vision*, 137-154.
- Wang C., et al. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*.
- Wang H., et al. (s.f.). Pose-guided part matching network via striking and reweighting for occluded person re-identification. *Image and Vision Computing*.
- Wang K., et al. (2021). D2C-Net: A Dual-branch, Dual-guidance and Cross-refine Network for Camouflaged Object Detection. *IEEE Transactions on Industrial Electronics*.
- Wang K., Peng X., Yang J., Meng D. y Qiao Y. (2020). Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition. *IEEE Transactions on Image Processing*, 29, 4057-4069.
- Wang L., et al. (2021). Giant Panda Identification. *IEEE Transactions on Image Processing*, 2837-2849.
- Wang Q., et al. (2020). Overview of deep-learning based methods for salient object detection in videos. *Pattern Recognition*, 107340.
- Wang S., et al. (2020). Visual Context Aware Ship Detector for High-Resolution SAR Imagery. *2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (págs. 1778-1781). IEEE.
- Wang W., et al. (2021). Salient Object Detection in the Deep Learning Era: An In-depth Survey. *Transactions on Pattern Analysis and Machine Intelligence*.
- Wang Y., et al. (2020). A Camouflaged Object Detection Model Based on Deep Learning. *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIS)* (págs. 150-153). IEEE.
- Wang Y., et al. (2022). Detecting Occluded and Dense Trees in Urban Terrestrial Views with High-quality Tree Detection Dataset. *IEEE Transactions on Geoscience and Remote Sensing*.
- Wang K., et al. (2020). Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition. *IEEE Transactions on Image Processing*, 4057-4069.
- Wei W., et al. (2019). Occluded Pedestrian Detection Based on Depth Vision Significance in Biomimetic Binocular. *IEEE Sensors Journal*, 11469-11474.
- Wei X., et al. (2020). Pedestrian detection in underground mines via parallel feature transfer network. *Pattern Recognition*, 107195.
- Wu C. y Ding J. (2018). Occluded face recognition using low-rank regression with generalized gradient direction. *Pattern Recognition*, 256-268.
- Wu D., et al. (2020). Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environment. *Computers and Electronics in Agriculture*, 105742.
- Wu T. (2020). A Supernova Detection Implementation based on Faster R-CNN. *2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)* (págs. 390-393). IEEE.
- Wu Y., et al. (2020). Real-time traffic sign detection and classification towards real traffic scene. *Multimedia Tools and Applications*, 18201-18219.
- Xiang L., et al. (2023). DINF: Dynamic Instance Noise Filter for Occluded Pedestrian Detection. *arXiv preprint arXiv:2301.05565*.

- Xiao B., et al. (2021). A vision-based method for automatic tracking of construction machines at nighttime based on deep learning illumination enhancement. *Automation in Construction*, 103721.
- Xiao Y., et al. (2020). A review of object detection based on deep learning. *Multimedia Tools and Applications*, 23729-23791.
- Xie H., et al. (2019). Context-aware pedestrian detection especially for small-sized instances with Deconvolution Integrated Faster RCNN (DIF R-CNN). *Applied Intelligence*, 1200-1211.
- Xie J., et al. (2020). Mask-Guided Attention Network and Occlusion-Sensitive Hard Example Mining for Occluded Pedestrian Detection. *IEEE Transactions on Image Processing*, 3872-3884.
- Xie S., et al. (2020). Diverse receptive field network with context aggregation for fast object detection. *Journal of Visual Communication and Image Representation*, 102770.
- Xu S., Gu J., et al. (2023). DKTNet: Dual-Key Transformer Network for small object detection. *Neurocomputing*, 29-41.
- Xu S., Zhang M., et al. (2023). A systematic review and analysis of deep learning-based underwater object detection. *Neurocomputing*, 204-232.
- Xu X., et al. (2021). Boundary guidance network for camouflage object detection. *Image and Vision Computing*, 104283.
- Xu Y., et al. (2018). Rapid Pedestrian Detection Based on Deep Omega-Shape Features with Partial Occlusion Handling. *Neural Processing Letters*, 923-937.
- Yan Z., et al. (2021). Detection-Oriented Backbone Trained from Near Scratch and Local Feature Refinement for Small Object Detection. *Neural Processing Letters*, 1921-1943.
- Yilmaz B. y Karşligil M. (2020). Detection Of Airplane And Airplane Parts From Security Camera Images with Deep Learning. *2020 28th Signal Processing and Communications Applications Conference (SIU)* (págs. 1-4). IEEE.
- Yuan J., et al. (2020). Gated CNN: Integrating multi-scale feature layers for object detection. *Pattern Recognition*, 107131.
- Yuan Y., Xiong Z. y Wang W. (2019). VSSA-NET: Vertical Spatial Sequence Attention Network for Traffic Sign Detection. *IEEE Transactions on Image Processing*, 3423-2434.
- Yudin D., et al. (2019). Detection of Big Animals on Images with Road Scenes using Deep Learning. *2019 International Conference on Artificial Intelligence Applications and Innovations* (págs. 100-103). IEEE.
- Zardoua Y., et al. (2021). A Horizon Detection Algorithm for Maritime Surveillance. *arXiv preprint*.
- Zhang B., et al. (2010). Local Derivative Pattern Versus Local Binary Pattern: Face Recognition With High-Order Local Pattern Descriptor. *IEEE Transactions on Image Processing*, 533-544.
- Zhang C., et al. (2022). Camouflaged object detection via Neighbor Connection and Hierarchical Information Transfer. *Computer Vision and image Understanding*, 103450.
- Zhang G., et al. (2004). Boosting Local Binary Pattern (LBP)-Based Face Recognition. *Chinese Conference on Biometric Recognition*, (págs. 179-186).
- Zhang G., et al. (2019). CAD-Net: A Context-Aware Detection Network for objects in Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 10015-10024.
- Zhang H. y Hong X. (2019). Recent progress on object detection: a brief review. *Multimedia Tools and Applications*, 27809-27847.
- Zhang H., et al. (2021). An empirical study of multi-scale object detection in high resolution UAV images. *Neurocomputing*, 173-182.

- Zhang K., et al. (2021). Semantic Context-Aware Network for Multiscale Object Detection in Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, 1-5.
- Zhang L., et al. (2020). A Multistage Refinement Network for Salient Object Detection. *IEEE Transactions on image Processing*, 3534-3545.
- Zhang P., et al. (2020). Non-rigid object tracking via deep multi-scale spatial-temporal saliency maps. *Pattern Recognition*, 107130.
- Zhang Q., et al. (2020). Dense Attention Fluid Network for Salient Object Detection in Optical Remote Sensing Images. *IEEE Transactions on Image Processing*, 1305-1317.
- Zhang S., et al. (2020). Asymmetric multi-stage CNNs for small-scale pedestrian detection. *Neurocomputing*, 12-26.
- Zhang T., et al. (2020). Omni-supervised joint detection and pose estimation for wild animals. *Pattern Recognition Letters*, 84-90.
- Zhang W., et al. (2021). Global context aware RCNN for object detection. *Neural Computing and Applications*, 11627-11639.
- Zhang X., et al. (2022). PSNet: Perspective-sensitive convolutional network for object detection. *Neurocomputing*, 384-395.
- Zhang X., Liu Y., Huo C., Xu N., Wang L. y Pan C. (2022). PSNet: Perspective-sensitive convolutional network for object detection. *Neurocomputing*, 384-395.
- Zhang Y., et al. (2019). Detecting small faces in the wild based on generative adversarial network and contextual information. *Pattern Recognition*, 74-86.
- Zhao M., et al. (2021). An occlusion-resistan circle detector using inscribed triangles. *Pattern Recognition*, 107588.
- Zheng H., et al. (2020). Feature enhacnement for multi-scale object detection. *Neural Processing Letters*, 1907-1919.
- Zheng Q. y Chen Y. (2021). Interactive multi-scale feature representation enhancement for small object detection. *Image and Vision Computing*, 104128.
- Zheng W., et al. (2020). A novel approach inspired by optic nerve characteristics for few-shot occluded face recognition. *Neurocomputing*, 25-41.
- Zheng Y., et al. (2019). Detection of People With Camouflage Pattern Via Dense Deconvolution Network. *IEEE Signal Processing Letters*, 29-33.
- Zhou C. y Yuan J. (2019). Multi Pattern Recognition-label learning of part detectors for occluded pedestrian detection. 99-111.
- Zhou C. y Yuan J. (2020). Occlusion Pattern Discovery for Object Detection and Occlusion Reasoning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2067-2080.
- Zhou K., et al. (2020). Improving Multispectral Pedestrian Detection by Addressing Modality Imbalance Problems. *European Conference on Computer Vision (ECCV)*, (págs. 787-803).
- Zhou S., et al. (2019). Discriminative Feature Learning With Foreground Attention for Person Re-Identification. *IEEE Transactions on Image Processing*, 4671-4684.
- Zhou S., et al. (2020). Depth occlusion perception feature analysis for person re-identification. *Pattern Recognition Letters*, 617-623.
- Zhou T., et al. (2021). RGB-D salient object detection: A survey. *Computational Visual Media*, 37-69.
- Zhou X. y Zhang L. (2022). SA-FPN: An effective feature pyramid network for crowded human detection. *Applied Intelligence*.

Zhu Y., et al. (2019). Attention CoupleNet: Fully Convolutional Attention Coupling Network for Object Detection. *IEEE Transactions on Image Processing*, 113-126.

Zhuge M., et al. (2022). CubeNet: X-shape connection for camouflaged object detection. *Pattern Recognition*, 108644.

Zou Z., et al. (2023). Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, 257-276.

Zuo J., et al. (2018). Person Re-Identification. *2018 IEEE International Conference on Multimedia and Expo (ICME)* (págs. 1-6). IEEE.



Esta obra está bajo una licencia de Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Sistema de Alerta Temprana epidemiológica animal

Animal Epidemiological Early Warning System

Edgar Montiel Cruz¹

Felipe Hernández González¹

Alejandro Luis Collantes Chávez-Costa¹

¹Universidad Autónoma de Quintana Roo

Resumen En la isla de Cozumel, México, se encuentran especies endémicas y ferales, mismas que se encuentran amenazadas por sus reducidos tamaños poblacionales y por enfermedades. Para el manejo de una problemática que se pudiera presentar y toma de decisiones, se propone un prototipo de sistema de alerta temprana epidemiológica animal, construida sobre una plataforma Web bajo el modelo cliente-servidor. Se siguió la metodología del ciclo de vida del software junto con el modelo incremental. Para la definición de los requerimientos se utilizó el Lenguaje Unificado de Modelado (Unified Model Language) integrando el desarrollo guiado por pruebas (Test Driven Development). Para la codificación se utilizó el lenguaje de programación JavaScript y MySQL como motor de base de datos. La implementación de la interfaz de usuario se realizó con el apoyo de las librerías Vue.js y Vuetify.js. Los módulos del sistema comprenden el catálogo de especies, captura de casos que incluyen datos de georeferencia y consultas de través de un mapeo de casos y un módulo de alertas automáticas based en el uso de semáforos similar al del SARS-COV 19.

Palabras clave: Sistema de Alerta temprana, Desarrollo Web.

Abstract: Endemic and feral species are found on the island of Cozumel, Mexico, which are threatened by their small population sizes and diseases. For the management of a problem that could arise and decision-making, a prototype of an animal epidemiological early warning system is proposed, built on a Web platform under the client-server model. The software life cycle methodology was followed following the incremental model. For the definition of the requirements, the Unified Model Language was used, integrating Test Driven Development. For coding, the programming language JavaScript and MySQL were used as the database engine. The implementation of the user interface was carried out with the support of the Vue.js and Vuetify.js libraries. The modules of the system include the catalog of species, capture of cases that include georeference data and queries through a case mapping and eaerly warning module which uses a traffic light system very similar SARS-COV 19.

Keywords: Early Warning System, Web Development

1 Introducción

La isla de Cozumel ubicada en el estado de Quintana Roo, México, es el primer destino de arribo de turistas de crucero del país. Tan sólo en el primer semestre de 2019 recibió 2,406,908 (Sectur, 2019) por esta vía, sin contar otros medios como el aéreo o cruce de ferry de pasajeros o de vehículos terrestres. En las áreas de selva virgen se encuentran especies endémicas, como *Nasua nelson* (coatí), *Procyon pygmaeus* (mapache enano), y *Pecari tajacu nanus* (Pecarí de Collar) (Universidad Nacional Autónoma de México, 2012), mismas que se encuentran amenazadas por sus reducidos tamaños poblacionales, por lo que pueden ser consideradas vulnerables a eventos destructivos como enfermedades. Estas especies comparten con animales domésticos, enfermedades infectocontagiosas, como lo son

la rabia y el distemper canino. Al avanzar la mancha urbana, es posible que se produzcan contactos entre las mismas y por consiguiente un contagio puede ocurrir (Frausto, Ihl, Rojas, 2016). Para contribuir a la conservación de las especies endémicas de la isla de Cozumel, el presente proyecto tiene la finalidad de elaborar un prototipo de sistema de alerta temprana epidemiológica animal, que ayude a la toma de decisiones en el control de enfermedades que amenazan la fauna endémica de la Isla de Cozumel.

Un sistema de alerta temprana se define como *“el conjunto de herramientas, dispositivos de control, capacidades de gestión e instrumentos tecnológicos que las instituciones claves identifican para difundir la información de manera oportuna a las comunidades expuestas a un riesgo, y cuyo resultado son medidas de mitigación orientadas a reducir los efectos de los*

desastres naturales y las pérdidas económicas y de vidas, así como las lesiones” (Domínguez, Lozano, 2019).

2 Métodos o Metodología

2.1 Ciclo de vida del software

Para el desarrollo del sistema de alerta temprana se siguieron las siguientes etapas del ciclo de vida del software:

- Investigación Preliminar. Se tomó en cuenta el contexto de la situación o problemática a resolver, mediante entrevistas con los usuarios.
- Definición de los Requerimientos. Se identificaron las características que debe tener el nuevo sistema, la información que deben producir y las características operacionales. El modelado de los requerimientos se hizo mediante la definición de casos de uso utilizando la notación **UML** (Unified Modeling Language) el cual se basa en el Paradigma Orientado a Objetos(**POO**) (Bruegge, Dutoit, 2009). Esto se ilustra en la figura 1.

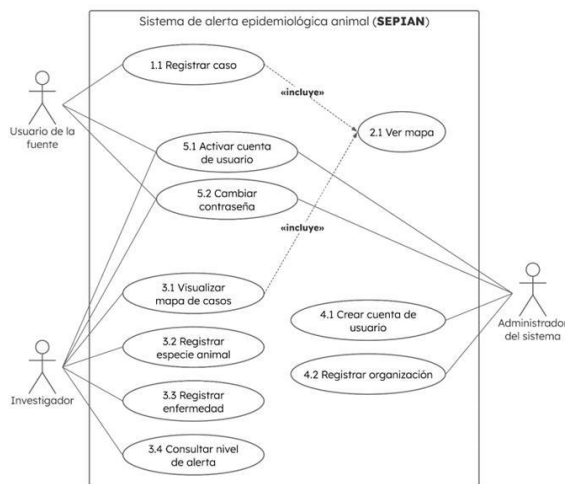


Figura 1. Casos de uso del sistema de alerta temprana animal.

- Definición de la Frontera del Problema. Se determinó con base a la definición de los requerimientos expresados en las entrevistas.
- Definición del Modelo Lógico de la Base de Datos. Se determinaron los detalles que establecen la forma en la que el sistema cumplirá con los requerimientos identificados durante la fase de análisis.
- Diseño Modular. En esta etapa se dividieron o fragmentaron los procesos del sistema en funciones separadas que deberán cumplir con ciertas condiciones.
- Diseño de la interfaz hombre-máquina. Se diseñó el conjunto de elementos con los que un usuario va a interactuar con la aplicación de software para llevar a cabo una actividad o

proceso. Comúnmente se denomina **GUI** (Graphic User Interface).

- Desarrollo del prototipo. Se transformó el diseño de los requerimientos en programas que cubran con la funcionalidad esperada. Una vez que se desarrolló el prototipo, se siguió el modelo de desarrollo de software incremental el cual consiste en lo siguiente: Después de contar con el desarrollo inicial de la arquitectura completa del sistema, se añaden incrementos y versiones parciales del mismo. Cada incremento tiene su propio ciclo de vida y se agregó funcionalidad adicional o mejorada sobre el sistema. Conforme se completó cada etapa, se verificó e integró la versión con las demás versiones ya completadas del sistema. Esto se puede ver en la figura 2.

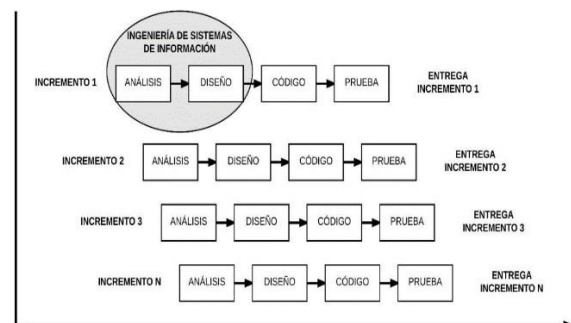


Figura 2. Modelo Incremental. Fuente: (Pressman,2001)

2.1.1 Desarrollo guiado por pruebas

Siguiendo esta práctica de programación, primero se definieron las pruebas o lo que el software tiene que hacer que es muy similar a la definición de casos de uso definidos en **RUP** (Rational Unified Process).

Posteriormente se elaboró el código fuente que pasó las pruebas satisfactoriamente. Una vez conseguido esto, se refactorizará el código escrito para ser parte del producto de software.

De esta forma el código es menos susceptible a fallas a la vez que se acelera el proceso de desarrollo. Para la codificación de las pruebas se utilizó la librería **Jest** que corre bajo **Node.js**.

Para la codificación del código fuente de todos los módulos incluyendo los pruebas, se empleó el **IDE** (Integrated Development Environment) Visual Studio Code en su última versión. Para el control de versiones lo más adecuado fue un repositorio en Gitlab cuyo manejo está integrado en Visual Studio Code.

2.1.2 Arquitectura del sistema

Se utilizó un patrón cliente-servidor y específicamente una arquitectura en capas. Se opta por el enfoque en capas a razón de su afinidad con el ciclo de vida del software y desarrollo incremental (Sommerville, 2011), haciéndolo adecuado para la utilización de una metodología de desarrollo ágil. Marston (2012) define las siguientes capas:

- Capa de presentación.
- Capa de lógica de negocio.
- Capa de acceso a datos.

La arquitectura del sistema por capas del sistema de alerta temprana animal se aprecia en la figura 3.

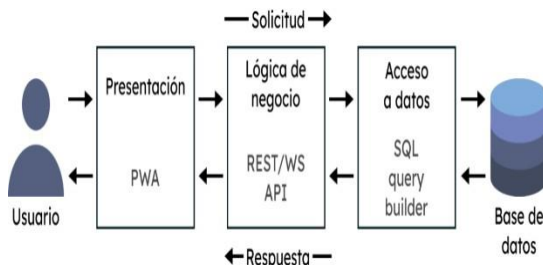


Figura 3. Arquitectura del sistema. Fuente: Elaboración propia

Implementación de las capas

En la capa de presentación será implementada la interfaz gráfica de usuario para el sistema. Se plantea el desarrollo de una Aplicación Web Progresiva (**PWA**, por sus siglas en inglés) la cual funciona independientemente del sistema operativo o dispositivo en el que se use. Una alternativa es el desarrollo de aplicaciones cliente nativas para cada sistema operativo requerido (Biørn-Hansen, Majchrzak, Grønli, 2017).

Se utilizó el lenguaje de programación JavaScript para el código fuente. En el cliente el motor de ejecución lo proveen los navegadores Web y en el servidor, el motor lo provee **Node.js**. Del lado del cliente se utilizaron las bibliotecas **Vue.js** y **Vuetify.js** que permiten la creación de una Interfaz Gráfica de Usuario de aspecto moderno. **Vue.js** permite la implementación de Progressive Web Applications, las cuales se comportan como una aplicación nativa de la plataforma dónde se ejecutan, ya sean computadoras de escritorio o dispositivos móviles (Lei, Ma, Tan, 2014). En el lado del

servidor se optó por **MySQL** como tecnología de base de datos, pudiendo ser reemplazada por cualquier otra que sea soportada en el entorno de **Node.js**.

También se empleó la librería **FeathersJS** para crear una **API REST** (Application Programming Interface basada en el protocolo **REST**) que proporcionará el acceso a los recursos de la base de datos a través del protocolo **HTTPS** (Hypertext Transfer Protocol Secure) e implementa la funcionalidad de autenticación de usuarios usando la tecnología **JSON Web Tokens**. En el servidor, se utilizó **Nginx** como proxy reverso para el acceso a la API, ya que proporciona balanceo de carga de ser requerido en el futuro. Adicionalmente el código del cliente requiere ser servido de forma estática mediante **HTTPS**, por lo que **Nginx** puede realizar esta función y servir a la **API** con un nombre de dominio y el código del cliente con otro.

Las tecnologías utilizadas se resumen en la figura 4.

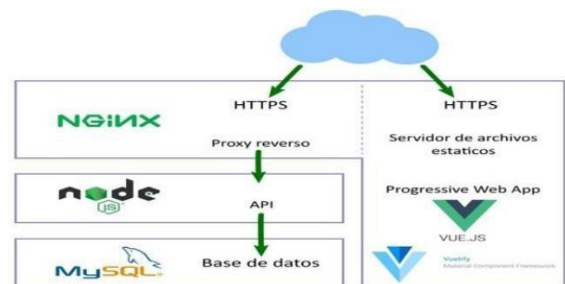


Figura 4. Tecnologías aplicadas a la arquitectura del sistema. Fuente: Elaboración propia

3 Resultados

3.1 Desarrollo del sistema

La interfaz del módulo principal del sistema se aprecia en la figura 5.

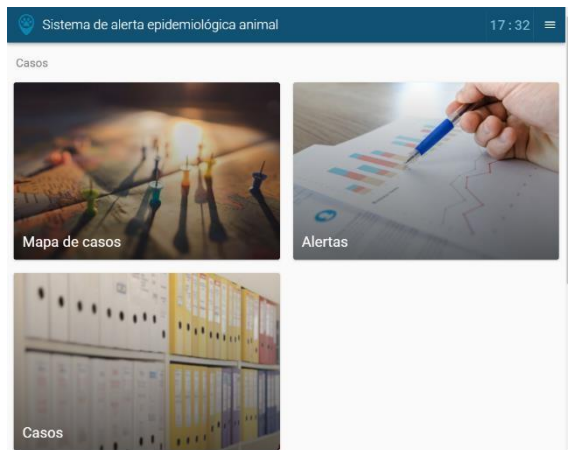


Figura 5. Módulos del sistema de alerta temprana animal.

Módulos principales: Casos, Mapa de Casos y Alertas.

La función del Módulo Casos es registrar la información de los casos de manera individual que incluye información georeferenciada como se puede apreciar en la figura 6.

Figura 6. Módulo de captura de casos.

La opción Mapa de Casos permite consultas de manera dinámica al especificar parámetros como la enfermedad o un rango de fechas. El resultado de una consulta se muestra en la figura 6.



Figura 7. Módulo de Consulta de Casos.

Las alertas del sistema, se definen por cada enfermedad registrada y se parametrizan con base a un nivel de incremento. Esto se aprecia en la figura 8.

Figura 8. Ejemplo de alertas automáticas.

De esta manera, se asocia este parámetro a un sistema de semáforo con códigos de color exactamente al que se utiliza por el gobierno de México para la contingencia sanitaria del SARS-COV-19: Verde (sin peligro), Amarillo (peligro bajo), Naranja (peligro moderado) y Rojo (peligro máximo).

Con esta información el sistema es capaz de generar alertas de manera automática. Esto se aprecia en la figura 9.

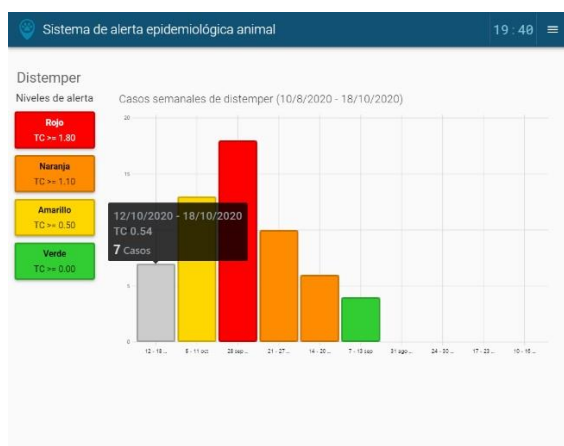


Figura 9. Ejemplo de alertas automáticas por rango de fechas. Fuente: Elaboración propia.

3.2 Desarrollo guiado por pruebas

El desarrollo guiado por pruebas (TDD, acrónimo de Test-Driven Development) es un enfoque donde convergen el diseño pruebas y el código.

Los pasos son los siguientes:

- 1.- Se identifica una funcionalidad con base en las reglas de negocio.
- 2.- Se diseña y construye la prueba, la cual es totalmente automatizada. Una vez termina se procede a ejecutarla.
- 3.- Si la prueba se supera con éxito, se procede a integrar una nueva funcionalidad, es decir se inicia otra vez el paso 1.
- 4.- En caso de que la prueba falle, se tienen que hacer las correcciones a la misma para una correcta implementación refactorizando el código. Una vez completado este proceso se vuelve a correr la prueba.

La figura 10 resume el proceso TDD.

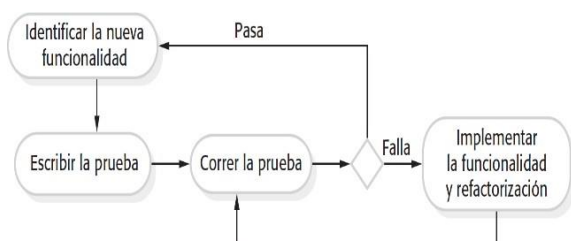


Figura 10. El desarrollo guiado por pruebas. Fuente: Basado en Dookhun, Nagowah, 2019.

El desarrollo dirigido por pruebas también funciona como una forma de documentar el desarrollo del código.

Un ejemplo del resultado del desarrollo guiado por pruebas se aprecia en la figura 11. Se utilizó el framework JUnit 5.

```

xrnoz@xrnoz-notebook: ~/Development/disease-tracking/back-end
xrnoz@xrnoz-notebook:~/Development/disease-tracking/back-end$ yarn test
yarn run v1.16.0
$ node --tls-min-v1.1 ./node_modules/jest/bin/jest.js
PASS tests/services/CasesService.test.js (7.106s)
PASS tests/services/SpeciesService.test.js (7.585s)
PASS tests/Service.test.js
PASS tests/Application.test.js
PASS tests/Database.test.js
PASS tests/services/UsersService.test.js (10.843s)

Test Suites: 6 passed, 6 total
Tests: 37 passed, 37 total
Snapshots: 0 total
Time: 12.985s
Ran all test suites.
Done in 15.46s.
xrnoz@xrnoz-notebook:~/Development/disease-tracking/back-end$

```

Figura 11. Resultado del Desarrollo Guiado por Pruebas para el Módulo de Casos.

4 Discusión

El haber seguido el ciclo de vida del software junto el desarrollo incremental permitió cumplir con el desarrollo del sistema. La implementación del desarrollo guiado por pruebas se complicó un poco en virtud que el equipo de trabajo no tenía experiencia ni con la metodología ni con las herramientas para llevarlas a cabo. Solo tenía experiencia en hacer pruebas de caja blanca para determinar los errores del software y corregirlos.

La definición de los parámetros de las alertas sufrió modificaciones en virtud que el usuario cambió los requerimientos iniciales que especificaban valores estáticos en lugar de valores que pudieran modificarse con el paso del tiempo según se necesitara con base al comportamiento de una enfermedad en específico. Se implementó un módulo para el administrador del sistema mismo que permite la creación de cuentas de usuario, que serán creadas para las personas que se encargarán de la captura de información.

Las consultas de casos por rango de fechas y el despliegue de estos en un mapa, fue una de las funcionalidades que el usuario consideró de mayor utilidad ya que le permitirá visualizar el comportamiento de una enfermedad y su área geográfica asociada. Permitirá el

establecimiento de cercos de seguridad sanitaria y acciones para control y erradicación de un brote.

5 Conclusiones

Se diseñó y construyó un Sistema de Alerta Temprana Animal que cumple con los requerimientos del usuario y tiene tres módulos completamente funcionales: Casos, Mapa de Casos y Alertas. El primero permite la captura de enfermedades nuevas en caso de que no se haya capturado con anterioridad. El segundo permite la generación de consultas por un rango de fechas y despliegue visual a través del mapeo de los casos con base a la información georeferenciada. El módulo de alertas permite la definición de las mismas con base al establecimiento de parámetros que pueden actualizarse a través del tiempo. Utiliza el sistema de semáforos como los empleados para el monitoreo del SARS-COV 19 para el despliegue visual de la información.

6 Trabajo a futuro

Es necesario que el sistema cuente con un módulo especial que permita explotar la información que se está registrando mediante la generación de estadísticas para complementar el módulo de consultas cuyos resultados de las mismas, se basan en un rango de fechas para generarlos. De esta manera se podrá dar seguimiento a la evolución de un brote que se presente, hacer comparaciones entre enfermedades, así como las bases para realizar un mapa de riesgos tomando la información georeferenciada de cada caso. También es necesario que el sistema genere gráficas para que para un mejor seguimiento del comportamiento y evolución de las enfermedades que se presenten.

Referencias bibliográficas

Dookhun, A. S., & Nagowah, L. (2019). Assessing the effectiveness of test-driven development and behavior-driven development in an industry setting. En 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 365-370). <https://doi.org/10.1109/ICCIKE47802.2019.900432>

Biørn-Hansen, A., Majchrzak, T. A., & Grønli, T.-M. (2017). Progressive Web Apps: The Possible Web-native Unifier for Mobile Development. Proceedings of the 13th International Conference on Web Information Systems and Technologies - Volume 1: WEBIST, 344–351. <https://doi.org/10.5220/0006353703440351>

Bruegge, B., & Dutoit, A. H. (2009). Object-oriented software engineering using UML, patterns, and Java (3a ed.). Pearson.

Dominguez, Lozano (2014). Recuperado el 12 de diciembre de 2019 de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0370-39082014000300007

Frausto, O., Ihl, T., & Rojas, J. (2016). Atlas de Riesgos de la Isla de Cozumel, México. Teoría Y Praxis, 12(Especial, Octubre 2016), 74–93.

Sommerville, I. (2011). Ingeniería de Software (9a ed.). Pearson Educación.

Lei, K., Ma, Y., & Tan, Z. (2014). Performance Comparison and Evaluation of Web Development Technologies in PHP, Python, and Node.js. 2014 IEEE 17th International Conference on Computational Science and Engineering, 661-668. <https://doi.org/10.1109/CSE.2014.142>

Marston, T. (2012). What is the 3-Tier Architecture? Recuperado el 30 de agosto de 2019 de <http://www.tonymarston.net/php-mysql/3-tier-architecture.html>



Esta obra está bajo una licencia de Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Control Neuro Proporcional-Derivativo para la formación Líder-Seguidor entrenado con el Filtro de Kalman Extendido

Neuro Proportional-Derivative Control for Leader-Follower Formation Trained with the Extended Kalman Filter

Dr. José de Jesús Hernández Barragán¹

Dra. Alma Yolanda Alanís García¹

Dr. Erasmo Gabriel Martínez Soltero¹

Dr. Jorge Daniel Ríos Arrañaga¹

¹Departamento de Innovación Basada en la Información y el Conocimiento, Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara 44430, México

Resumen: Este trabajo presenta un esquema control neuro adaptativo para la formación Líder-Seguidor de robots móviles diferenciales. El esquema está basado en un control neuro proporcional derivativo entrenado con filtro de Kalman extendido. El controlador propuesto ofrece un ajuste de ganancias adaptativo en línea, lo cual es ideal para lidiar con problemas de perturbaciones, ruido y dinámicas no modeladas. El esquema propuesto es implementado a nivel simulación y experimental usando robots Turtlebot3® y la plataforma Sistema Operativo Robótico (ROS). Además, el desempeño del controlador propuesto es comparado con un controlador para la formación Líder-Seguidor del estado del arte. Los resultados muestran que la propuesta de control tiene un mejor tiempo de convergencia, menor error de formación y menor error de seguimiento.

Palabras clave: Formación, Control PD Adaptativo, Filtro de Kalman Extendido, Robot móvil, Turtlebot3.

1. Introducción

La interacción entre un sistema robótico con otros sistemas robóticos o humanos de manera cooperativa para realizar una tarea es crucial para resolver problemas de vigilancia, exploración, reconocimiento, monitoreo ambiental, y manipulación cooperativa. Esta metodología es nombrada sistemas de robots de red (Sanfeliu, Hagita, & Saffiotti, 2008), (Kamel, Yu, & Zhang, 2020), (Wang, y otros, 2016). Realizar tareas con más de un equipo trae ventajas como: multitarea, tolerancia a fallas, rentabilidad, flexibilidad y distribución. Un sistema cooperativo multi agente puede estar compuesto por robots que operan coordinados para realizar una tarea. La técnica que garantiza la cooperación entre robots se denomina control cooperativo. El control de formación para múltiples agentes se considera un caso especial de control cooperativo (Kamel, Yu, & Zhang, 2020). El objetivo del control para la formación es mantener cierto patrón o forma durante la tarea de ejecución. Este patrón define una formación, que prácticamente son distancias relativas entre robots (Oh, Park, & Ahn, 2015). Existe una gran variedad de estrategias para el control de formación. Entre las más comunes se encuentran métodos basados en estructuras virtuales, esquemas de Líder-Seguidor, basados en grafos, y campos de potencial artificial (Kamel, Yu, & Zhang, 2020). En el esquema de Líder-Seguidor, uno de los robots se asigna como líder, mientras que los otros robots son seguidores (Wang, y otros, 2016). La ley de control se aplica a cada seguidor de manera independiente al líder. El propósito es que el robot seguidor mantenga una posición de formación con respecto al líder. En este trabajo, la estrategia de control de formación propuesta se enfoca en el esquema de Líder-Seguidor.

A continuación, mencionamos algunas de las estrategias del estado del arte sobre el control para la formación Líder-Seguidor. En (Lu, et al., 2019), se introduce un controlador que lidia con limitaciones de distancia en la comunicación entre robots móviles diferenciales. Luego en (Xuan-Mung & Hong, 2019), se propone una técnica de control robusta ante la presencia de perturbaciones, durante la tarea para la formación de drones. Después en (Hassan & Hammuda, 2020), se propone un esquema de control con base en observadores que permiten estimar el estado de los robots con alta precisión, lo que mejora la estrategia de control. Por otra parte, en (Dai, He, Chen, & Jin, 2020), se propone una estrategia de control distribuido utilizando solo información local entre los robots. Esta estrategia involucra un modelo dinámico para la formación. Finalmente, se propone un esquema de control para la formación de robots móviles tipo coche, especialmente en el caso del modelo Ackerman (Alfaro & Morán, 2020). Aunque estas estrategias demuestran su capacidad para lograr y mantener una formación, muchos de los trabajos reportan solo resultados de simulación. En una aplicación real, se deben tomar en consideración perturbaciones externas, ruido de medición y dinámicas no modeladas. El desempeño de estas estrategias se puede ver afectado en la práctica. Esto implica un ajuste de ganancias experimentales para mejorar el desempeño de las estrategias de control. Para resolver este

inconveniente, en este trabajo se plantea una estrategia de control adaptativo para la formación Líder-Seguidor.

Por otro lado, en la actualidad la presencia e importancia de la inteligencia artificial no se puede ignorar, ha permitido un avance en la sociedad en relativamente poco tiempo y sigue avanzando al encontrarse en áreas clave de la industria, la educación, la ciencia y el entretenimiento (Bryson, 2019). Día a día se proponen soluciones basadas en las diversas metodologías de inteligencia artificial, dentro de control de sistemas, sin duda a pesar de sus desventajas conocidas los controladores P, PD, y PID siguen siendo de los controladores más utilizados. Los controladores PID neuronales proponen metodologías para mejorar su desempeño del PID clásico. Existen técnicas de PID adaptativas entrenadas con base en gradientes (Hernandez-Alvarado, Garcia-Valdovinos, Salgado-Jimenez, Gómez-Espinosa, & Fonseca-Navarro, 2016), (Tang, Wang, Gu, & Gu, 2020), (Zhong, Zhu, Zhao, Han, & Zhang, 2020). Aunque se ha demostrado que los controladores PID entrenados con el Filtro de Kalman Extendido muestran superioridad con respecto a tasas de aprendizaje y tiempos de convergencias más rápidos (Hernandez-Barragan, et al., 2020), (Hernandez-Barragan, et al., 2021). En este artículo, la estrategia de control adaptativo para la formación Líder-Seguidor se basa en un controlador PD de una sola neurona, entrenada con el filtro de Kalman extendido.

El aporte de este trabajo está en el diseño de un esquema control neuro adaptativo para la formación Líder-Seguidor de robots móviles diferenciales. El esquema de control propuesto se basa en un controlador neuro proporcional derivativo (NPD, por sus siglas en inglés, *Neural Proportional-Derivative*) que es entrenado con el filtro de Kalman extendido. La aportación principal del controlador NPD es evitar el ajuste de ganancias del controlador, ante perturbaciones exteriores, ruido y dinámicas no modeladas, especialmente para el caso de una aplicación práctica. Además, el controlador NPD se inspira y compara con un controlador del estado del arte para la formación Líder-Seguidor, tanto en pruebas de simulación como en pruebas de experimentación.

El trabajo está organizado de la siguiente manera: en la sección 2, se presenta el modelo cinemático del robot móvil diferencial, la cinemática para la formación Líder-Seguidor y la estrategia de control que toma como base el diseño del controlador propuesto NPD. Después, en la sección 3 se describe el diseño del controlador NPD, desde el diseño de la neurona y su entrenamiento, hasta la descripción del controlador adaptativo para la formación Líder-Seguidor. En la sección 4 se presentan los resultados de la simulación y la experimentación. La discusión sobre los resultados se muestra en la sección 5. Finalmente, la conclusión y el trabajo futuro se presenta en la sección 6.

2. Preliminares

2.1 Modelo cinemático del robot móvil

Los robots de tipo diferencial son equipos robóticos ampliamente utilizados en la industria junto con la metodología de sistema de robots en red en las áreas mencionadas en el párrafo anterior, estos equipos se caracterizan por contar con dos ruedas que son motorizadas de manera independiente, su tracción se ve afectado por el tipo de ruedas y terreno, por lo que se puede contar con diferentes modelos de robots en esta clasificación, como los modelo tipo oruga y el de dos ruedas (Siegwart, Nourbakhsh, & Scaramuzza, 2011), (Ben-Ari & Mondada, 2017).

El diagrama del robot móvil diferencial considerado en este trabajo se muestra en la Figura 1, donde la posición (x, y) y la orientación θ del robot están expresados en el marco de referencia global $\{o\}$. El marco de referencia local $\{b\}$ esta adherido a la plataforma móvil, justo en el centro de las ruedas. Además, v es una velocidad lineal y w una velocidad angular, ambas definidas en el marco de referencia local $\{b\}$.

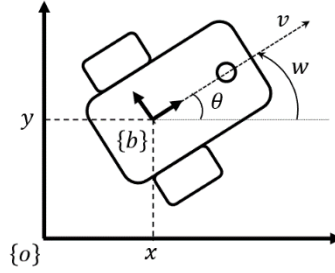


Figura 1: Diagrama del robot móvil diferencial.

El modelo cinemático del robot móvil está definido en (1), donde las velocidades $(\dot{x}, \dot{y}, \dot{\theta})$ describen el movimiento del robot con respecto al marco de referencia global $\{o\}$. Este modelo expresa la cinemática del robot monociclo donde las velocidades (v, w) definen las variables de control (Klancar, Zdesar, Blazic, & Skrjanc, 2017). En este trabajo, se considera controlar el robot móvil diferencial con las variables de control del modelo monociclo.

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & 0 \\ \sin(\theta) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} \quad (1)$$

2.2 Modelo cinemático para la formación Líder-Seguidor

El sistema de formación utilizado en este trabajo está descrito en la Figura 2: Esquema de formación Líder-Seguidor. La formación consiste en un robot móvil líder i y un robot móvil seguidor j . Las velocidades (v_i, w_i) y (v_j, w_j) expresan las velocidades lineales y angulares para cada robot, medidas en los marcos de referencia locales $\{i\}$ y $\{j\}$, respectivamente. Así mismo, se define la posición $\mathbf{r}_i = [x_i \ y_i]^T$ y orientación θ_i del robot líder con respecto al marco de referencia global. De manera similar, definimos la posición $\mathbf{r}_j = [x_j \ y_j]^T$ y orientación θ_j del robot seguidor.

La posición $\mathbf{r}_{ij} = [x_{ij} \ y_{ij}]^T$ está medida desde el marco de referencia del líder $\{i\}$ y se puede expresar como en (2).

$$\mathbf{r}_{ij} = \begin{bmatrix} \cos(\theta_i) & \sin(\theta_i) \\ -\sin(\theta_i) & \cos(\theta_i) \end{bmatrix} (\mathbf{r}_j - \mathbf{r}_i) \quad (2)$$

Además, la orientación θ_{ij} del robot seguidor se puede expresar desde el marco de referencia $\{i\}$ como $\theta_{ij} = \theta_j - \theta_i$.

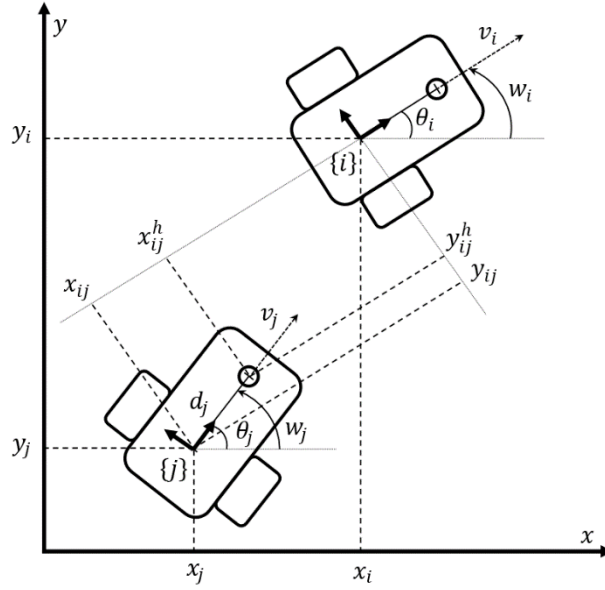


Figura 2: Esquema de formación Líder-Seguidor.

Por propósitos del control de formación, se define un punto $\mathbf{r}_{ij}^h = [x_{ij}^h \ y_{ij}^h]^T$ localizado a una distancia d_j desde el marco de referencia local $\{j\}$ del seguidor. Este punto se puede calcular como se expresa en (3).

$$\mathbf{r}_{ij}^h = \mathbf{r}_{ij} + d_j \begin{bmatrix} \cos(\theta_{ij}) \\ \sin(\theta_{ij}) \end{bmatrix} \quad (3)$$

Finalmente, derivando $\mathbf{r}_{ij}^h$ y la orientación θ_{ij} con respecto al tiempo se obtiene el modelo cinemático

(4), donde $\mathbf{G}_{ij}$ se define en (5) (Liang, et al., 2016) (Liang, Wang, Liu, Liu, & Chen, 2020).

$$\begin{bmatrix} \dot{x}_{ij}^h \\ \dot{y}_{ij}^h \end{bmatrix} = \mathbf{G}_{ij} \begin{bmatrix} v_j \\ w_j \end{bmatrix} + \begin{bmatrix} -v_i \\ 0 \end{bmatrix} + w_i \begin{bmatrix} y_{ij}^h \\ -x_{ij}^h \end{bmatrix} \quad (4)$$

$$\dot{\theta}_{ij} = w_j - w_i$$

$$\mathbf{G}_{ij} = \begin{bmatrix} \cos(\theta_{ij}) & -d_j \sin(\theta_{ij}) \\ \sin(\theta_{ij}) & d_j \cos(\theta_{ij}) \end{bmatrix} \quad (5)$$

Para más detalles sobre la cinemática para la formación Líder-Seguidor, se recomienda consultar (Liang, Wang, Liu, Liu, & Chen, 2020).

2.3 Control para la formación Líder-Seguidor

En (Liang, Wang, Liu, Liu, & Chen, 2020), se propone el diseño de un controlador para la formación Líder-Seguidor. En este trabajo, se hace referencia a esta estrategia como el controlador de formación [Liang 2020]. En esta estrategia, se define una posición de referencia $\mathbf{r}_{ij}^* = [x_{ij}^* \ y_{ij}^*]^T$ relativa al marco de referencia del líder $\{i\}$. En otras palabras, el vector $\mathbf{r}_{ij}^*$ define una formación deseada. El propósito del controlador es calcular las velocidades (v_j, w_j) del robot seguidor para llevar la posición actual $\mathbf{r}_{ij}^h$ a la posición de referencia $\mathbf{r}_{ij}^*$.

La estrategia de control de formación [Liang 2020] define como en (6), donde $\mathbf{K}_{ij}$ es una matriz simétrica definida positiva y $\Delta \mathbf{r}_{ij}^* = \mathbf{r}_{ij}^h - \mathbf{r}_{ij}^*$ es el error de formación. La matriz $\mathbf{K}_{ij}$ define las ganancias del control de formación. Para poder implementar el controlador, se asume que las posiciones y orientaciones relativas de cada robot son accesibles. También se asume que las velocidades del robot líder (v_i, w_i) son medibles.

$$\begin{bmatrix} v_j \\ w_j \end{bmatrix} = \mathbf{G}_{ij}^{-1} \left\{ -\mathbf{K}_{ij} \Delta \mathbf{r}_{ij}^* - \begin{bmatrix} -v_i \\ 0 \end{bmatrix} - w_i \begin{bmatrix} y_{ij}^h \\ -x_{ij}^h \end{bmatrix} \right\} \quad (6)$$

En el esquema de control (6), el cálculo de las velocidades del robot (v_j, w_j) , no solo dependen de los ajustes de la matriz de ganancias $\mathbf{K}_{ij}$, también depende de las velocidades del robot líder (v_i, w_i) . Es decir, valores altos para las velocidades (v_i, w_i) provocan valores altos para (v_j, w_j) . En este caso, las ganancias de $\mathbf{K}_{ij}$ no deberían de ser grandes para no calcular velocidades excesivas para el robot seguidor. Por otra parte, valores bajos para (v_i, w_i) provocan valores bajos para (v_j, w_j) . En este otro caso, las ganancias de $\mathbf{K}_{ij}$ deberían de ser mas grandes para llevar los errores $\Delta \mathbf{r}_{ij}^*$ a un vecindario cercano a 0. La desventaja de esta estrategia de control es que la ganancia $\mathbf{K}_{ij}$ es fija.

3. Descripción del Controlador Neuro Proporcional-Derivativo para la formación Líder-Seguidor

3.1 Controlador Neuro Proporcional-Derivativo

El esquema del controlador NPD se muestra en la Figura 3. El controlador NPD está compuesto por una neurona de dos entradas y una salida.

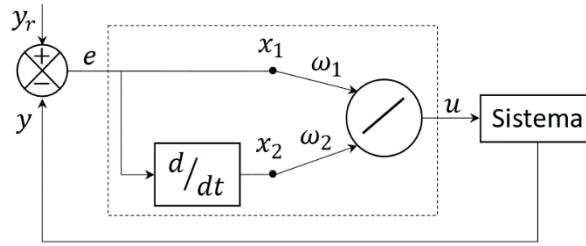


Figura 3: Controlador Neuro Proporcional-Derivativo.

El error $e(k)$ es la diferencia (7) entre una referencia $y_r(k)$ y la salida del sistema $y(k)$ en el instante de tiempo k . Las entradas $x_1(k)$ y $x_2(k)$ están definidas como el error (8) y la derivada del error (9), respectivamente en tiempo discreto (Hernandez-Barragan, et al., 2021). Los pesos de la neurona representan las ganancias del control, donde $\omega_1(k)$ y $\omega_2(k)$ son la ganancia proporcional y la ganancia derivativa, respectivamente. Estos pesos se ajustan en línea utilizando el algoritmo FKE. Finalmente, la salida $u(k)$ representa la acción de control (10).

$$e(k) = y_r(k) - y(k) \quad (7)$$

$$x_1(k) = e(k) \quad (8)$$

$$x_2(k) = e(k) - e(k-1) \quad (9)$$

$$u(k) = \omega_1(k)x_1(k) + \omega_2(k)x_2(k) \quad (10)$$

3.2 Algoritmo de entrenamiento con base en el Filtro de Kalman Extendido

El algoritmo de entrenamiento con base en el Filtro de Kalman Extendido (FKE) proporciona tasas de aprendizaje y tiempo de convergencia más rápidos que los algoritmos de entrenamiento con base en gradientes (Hernandez-Barragan, et al., 2020), (Hernandez-Barragan, et al., 2021). Además, el uso del algoritmo FKE es ideal para aplicaciones en línea y en tiempo real.

El FKE tiene como objetivo encontrar los pesos óptimos de la neurona que minimizan el error de predicción (Sanchez & Alanis, Redes neuronales: conceptos fundamentales y aplicaciones a control automático., 2006). Los pesos de la neurona se ajustan utilizando las reglas (11), (12) y (13) de actualización en el instante de tiempo k .

$$\mathbf{K}(k) = \mathbf{P}(k)\mathbf{H}(k)[\mathbf{R}(k) + \mathbf{H}^T(k)\mathbf{P}(k)\mathbf{H}(k)]^{-1} \quad (11)$$

$$\boldsymbol{\omega}(k+1) = \boldsymbol{\omega}(k) + \eta\mathbf{K}(k)e(k) \quad (12)$$

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \mathbf{K}(k)\mathbf{H}^T(k)\mathbf{P}(k) + \mathbf{Q}(k) \quad (13)$$

donde

- $\boldsymbol{\omega}(k)$ es el estado que representa los pesos de la neurona $\boldsymbol{\omega}(k) = [\omega_1(k) \ \omega_2(k)]^T$.
- $\mathbf{K}(k) \in \mathbb{R}^{2 \times 1}$ es el vector de ganancias de Kalman.
- $\mathbf{P}(k) \in \mathbb{R}^{2 \times 2}$ es la matriz de covarianza de error de predicción.
- $\mathbf{Q}(k) \in \mathbb{R}^{2 \times 2}$ es la matriz de covarianza del ruido del proceso.
- $\mathbf{R}(k) \in \mathbb{R}$ representa el error de covarianza de medición.
- $\mathbf{H}(k) \in \mathbb{R}^{2 \times 1}$ contiene las derivadas parciales de la salida de la red $u(k)$ con respecto a cada uno de los pesos.
- η es un factor de aprendizaje.

Los parámetros de Kalman η , $\mathbf{P}$, $\mathbf{Q}$ y $\mathbf{R}$ se ajustan de manera experimental. El factor de aprendizaje η se ajusta con la intención de minimizar el error $e(k)$. Las matrices $\mathbf{P}$ y $\mathbf{Q}$ se inicializan con valores en la diagonal principal, mientras que $\mathbf{R}$ es un escalar. La matriz $\mathbf{Q}$ se ajusta para lidiar con el ruido del proceso y $\mathbf{R}$ con el ruido de medición. Finalmente, el vector $\mathbf{H}(k)$ está definido como (14).

$$\mathbf{H}(k) = \begin{bmatrix} \frac{\partial u(k)}{\partial \omega_1(k)} \\ \frac{\partial u(k)}{\partial \omega_2(k)} \end{bmatrix} = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \quad (14)$$

Mas detalles sobre el entrenamiento de redes neuronales con base en el algoritmo FKE se pueden encontrar en (Sanchez, Alanís, & Loukianov, Discrete-time high order neural control, 2008).

Durante el entrenamiento de la neurona, los pesos se podrían ajustar con valores negativos. Así mismo, los pesos pueden crecer sin límite. En este trabajo, recomendamos mantener los pesos de la neurona acotados. Es importante asegurar (15), donde ω_1^{max} y ω_2^{max} definen los límites máximos de ajuste de la parte proporcional y derivativa, respectivamente.

$$0 \leq \omega_1 \leq \omega_1^{max}, \quad 0 \leq \omega_2 \leq \omega_2^{max} \quad (15)$$

3.3 Controlador Neuro Proporcional-Derivativo para la formación Líder-Seguidor

En este trabajo, se propone utilizar el enfoque de control NPD para dotar al controlador (6) con la capacidad de adaptar sus ganancias en línea. Se pretende sustituir solo el termino $\mathbf{K}_{ij}\Delta\mathbf{r}_{ij}^*$ por dos controladores NPD. Las demás ecuaciones del controlador permanecen sin modificación.

Dada la formación deseada $\mathbf{r}_{ij}^*$ y la posición actual del robot seguidor $\mathbf{r}_{ij}^h$, se puede utilizar el error de formación $\Delta\mathbf{r}_{ij}^* = \mathbf{r}_{ij}^h - \mathbf{r}_{ij}^* = \mathbf{e}$ como un error en el instante de tiempo k definido como $\mathbf{e}(k) = [e_x(k) \ e_y(k)]^T$. Dado que el vector de error $\mathbf{e}(k)$ está compuesto por dos errores $e_x(k)$ y $e_y(k)$, es necesario utilizar dos controladores NPD. La salida de cada controlador adaptativo define una acción de control $\mathbf{u}(k)$ tal que $\mathbf{u}(k) = [u_x(k) \ u_y(k)]^T$.

La estrategia de control se puede reescribir como (16).

$$\begin{bmatrix} v_j \\ w_j \end{bmatrix} = \mathbf{G}_{ij}^{-1} \left\{ -\mathbf{u}(k) - \begin{bmatrix} -v_i \\ 0 \end{bmatrix} - w_i \begin{bmatrix} y_{ij}^h \\ -x_{ij}^h \end{bmatrix} \right\} \quad (16)$$

Durante las siguientes secciones, el desempeño del controlador NPD propuesto (16) se comparará con el controlador [Liang 2020] (6).

4. Resultados

En esta sección se muestra el desempeño del controlador NPD propuesto para la formación Líder-Seguidor con base en simulaciones y experimentos. También consideramos incluir una comparación del controlador NPD propuesto (16) con el controlador [Liang 2020] (6). Los detalles para las pruebas de simulación y experimentación se muestran a continuación.

Se incluye una prueba de control en nivel simulación y otra prueba a nivel experimental. En cada prueba, el robot móvil líder i sigue una trayectoria variante en el tiempo de manera independiente al robot móvil seguidor. El propósito del control de formación es calcular la acción de control (v_j, w_j) para el robot seguidor j con la intención de que mantenga una formación deseada $\mathbf{r}_{ij}^*$ con respecto al robot líder i .

En cada prueba, la posición inicial del robot líder esta selecciona como $\mathbf{r}_i = [0 \ 0]^T$ y la orientación $\theta_i = 0$. Para el robot seguidor, su posición inicial es $\mathbf{r}_j = [-0.9 \ -0.3]^T$ con orientación $\theta_j = 0$. Además, se consideró utilizar $d_j = 0.08$ para el robot seguidor y para la formación deseada $\mathbf{r}_{ij}^* = [-0.6 \ 0]^T$.

Para la matriz de ganancias $\mathbf{K}_{ij}$ del controlador clásico, se utilizó (17).

$$\mathbf{K}_{ij} = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.3 \end{bmatrix} \quad (17)$$

En el caso del control NPD, el factor de aprendizaje se estableció como $\eta = 0.05$. Los pesos iniciales de la red se consideraron como $\boldsymbol{\omega} = [0.3 \ 0]^T$. Los pesos máximos para limitar las ganancias del control NPD se seleccionaron como $\omega_1^{max} = \omega_2^{max} = 2.0$. Estos parámetros se configuraron de manera experimental. Finalmente, parámetros $\mathbf{R}$, $\mathbf{P}$ y $\mathbf{Q}$ de Kalman se establecieron como (18), (19) y (20), respectivamente, de acuerdo con (Hernandez-Barragan, et al., 2020), (Hernandez-Barragan, et al., 2021).

$$\mathbf{R} = 0.001 \quad (18)$$

$$\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (19)$$

$$\mathbf{Q} = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix} \quad (20)$$

Para comparar el desempeño de las estrategias de control para la formación líder seguidor, se muestran los resultados de los errores de formación $\Delta \mathbf{r}_{ij}^* = \mathbf{r}_{ij}^h - \mathbf{r}_{ij}^* = [\Delta x_{ij} \quad \Delta y_{ij}]^T$ en gráficas.

También, se considera incluir gráficas para comparar el seguimiento de formación expresado desde el marco de referencia global. Es decir, las gráficas muestran la formación de referencia $\mathbf{r}_j^*$ y la posición actual $\mathbf{r}_j^h$. La posición $\mathbf{r}_j^h$ se puede expresar desde el marco de referencia global como (21).

$$\mathbf{r}_j^h = \begin{bmatrix} x_j^h \\ y_j^h \end{bmatrix} = \mathbf{r}_j + d_j \begin{bmatrix} \cos(\theta_j) \\ \sin(\theta_j) \end{bmatrix} \quad (21)$$

Además, la posición deseada $\mathbf{r}_{ij}^*$ se puede calcular desde el marco de referencia global con $\mathbf{r}_j^*$ definido como (22).

$$\mathbf{r}_j^* = \begin{bmatrix} x_j^* \\ y_j^* \end{bmatrix} = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i) \end{bmatrix} \mathbf{r}_{ij}^* + \mathbf{r}_i \quad (22)$$

Finalmente, la acción de control (v_j, w_j) de ambas estrategias también se considera reportar con graficas.

Las especificaciones de equipo de cómputo utilizado para las pruebas es procesador Intel Core i7®-4770 (Intel Core i7, es una marca registrada de Intel Corporation, EE. UU.) con CPU a 3.4 GHz y 16 GB de RAM. El sistema operativo Windows® 10 (Windows, es una marca registrada de Microsoft Corporation, EE. UU.) y el software Matlab® R2021a (Matlab, es una marca registrada de Mathworks, EE. UU.). Los Turtlebot3® corren bajo Ubuntu® 18.04 (Ubuntu, es una marca registrada de Canonical Ltd., Reino Unido) y ROS Kinetic (ROS, es una marca registrada de Open Robotics, EE.UU.).

4.1 Resultados de simulación

Para la prueba de simulación, se cuenta con todas las mediciones necesarias, tanto posiciones y orientaciones relativas y globales, así como las velocidades del robot móvil líder (v_i, w_i) . Los resultados de la prueba se muestran a continuación.

Los errores de formación $\Delta \mathbf{r}_{ij}^*$ para las estrategias de formación comparadas se muestran en la Figura 4. La Figura 4 (a) muestra que el controlador [Liang 2020] converge a un vecindario cercano a cero después de 15 segundos de simulación. También se puede observar que el error Δy_{ij} converge un poco más rápido que Δx_{ij} . Así mismo, la Figura 4 (b) muestra que la estrategia de control propuesta NPD converge antes de 5 segundos. Además, ambos errores Δx_{ij} y Δy_{ij} convergen de manera similar. Este efecto se debe notar en la grafica de seguimiento de formación. De acuerdo con estos resultados, el controlador NPD reporta un tiempo de convergencia menor comparado con el controlador [Liang 2020].

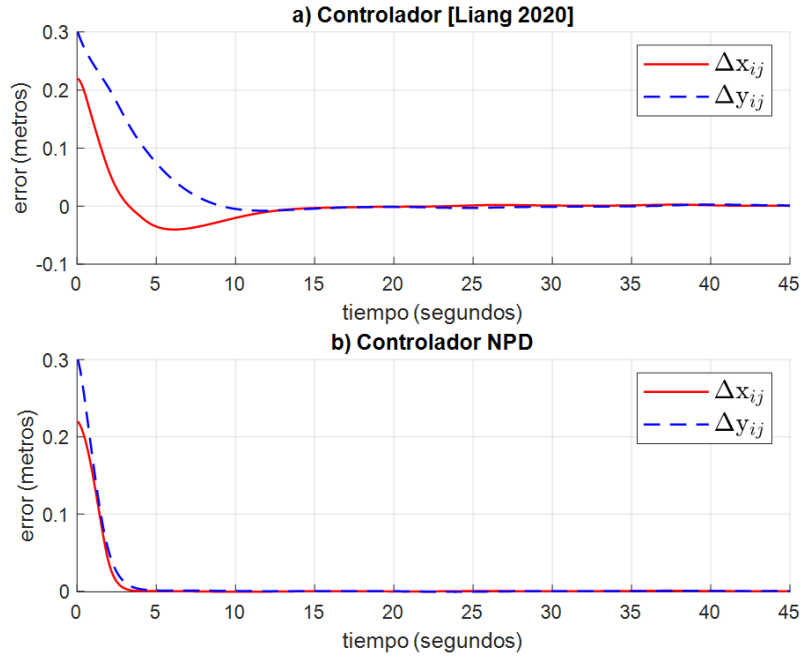


Figura 4: Errores de formación $\Delta \mathbf{r}_{ij}^*$. a) muestra los resultados de controlador [Liang 2020], mientras que b) muestra los resultados del control propuesto.

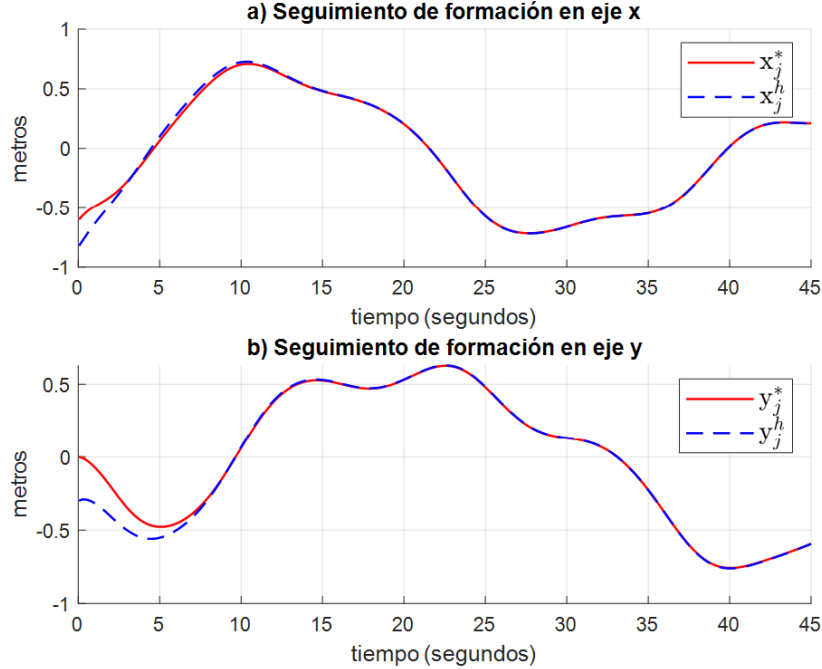


Figura 5: Seguimiento de formación del controlador [Liang 2020].

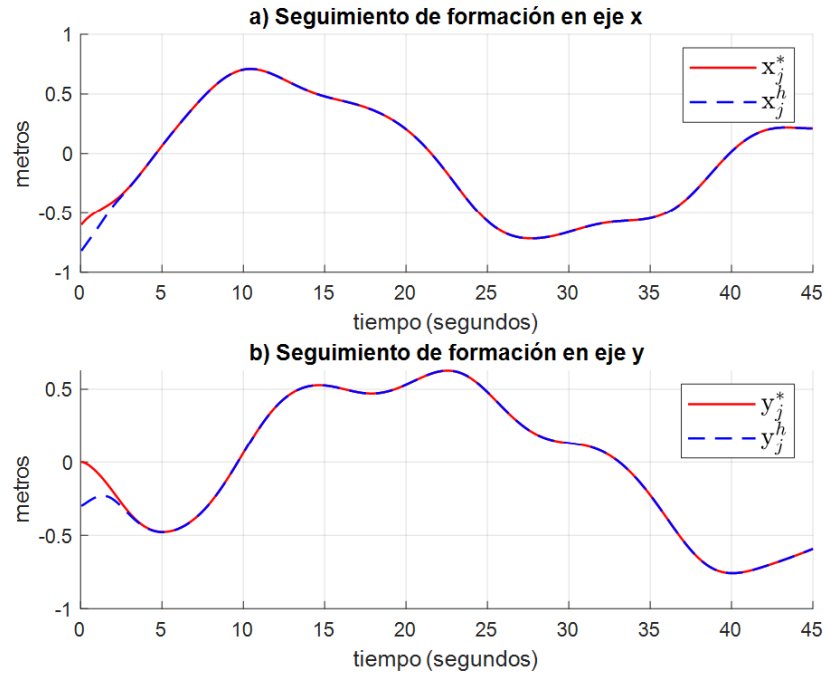


Figura 6: Seguimiento de formación del controlador propuesto NPD.

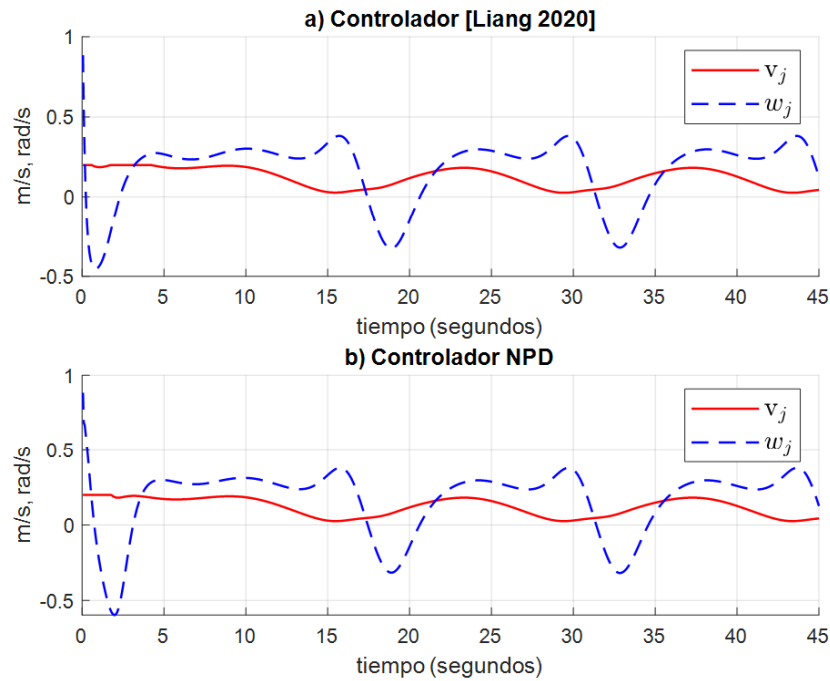


Figura 7: Acción de control para robot móvil seguidor j . a) muestra los resultados de controlador clásico, mientras que b) muestra los resultados del control propuesto. m: metros, rad: radianes, s: segundos.

La Figura 5 reporta los resultados del seguimiento de formación para el caso del controlador [Liang 2020]. Como se esperaba al analizar los errores de formación, después de los 15 segundos la referencia y la posición actual son similares. Además, se corrige más rápido el seguimiento en el eje y . Por otra parte, la Figura 6 muestra los resultados del seguimiento de formación para el caso del controlador NPD. Debido a la rápida convergencia del error, la posición actual alcanza la referencia antes de 5 segundos. Ambas estrategias de control demuestran un seguimiento de formación adecuado, sin embargo, el controlador propuesto NPD resuelve el seguimiento en menor tiempo.

Para terminar con los resultados de simulación, la acción de control para el robot móvil seguidor se reporta en la Figura 7. Las velocidades que se aplican al robot seguidor son similares para ambas estrategias de control. En el caso del controlador propuesto NPD de la Figura 7 (b), la velocidad angular w_j es menor a -0.5 rad/s. Al contrario, la velocidad angular w_j es mayor a -0.5 rad/s para el caso del controlador clásico de la Figura 7 (a). Esto indica que el controlador NPD calculó una acción de control mayor para corregir el seguimiento en menor tiempo. Después de los 10 segundos no se nota una diferencia significativa entre las velocidades de ambos controladores. Finalmente, las velocidades lineales y angulares reportadas son adecuadas para su aplicación en el Turtlebot3®.

4.2 Resultados de experimentación

Para la prueba de experimentación, se utilizaron dos robots móviles Turtlebot3®. La plataforma móvil Turtlebot3 es un robot diferencial para educación e investigación de bajo costo (ROBOTIS, 2022). Además, tiene muchas herramientas de software y hardware libre (ROBOTIS, 2022). Uno de los modelos de Turtlebot3® es el Waffle Pi. Entre sus componentes principales se encuentran: una tarjeta Raspberry Pi 3, dos actuadores XL430-W20, un módulo IMU con giroscopio y acelerómetro de 3 ejes, un LiDAR LDS-01, una tarjeta openCR (32-bit ARM Cortex M7) para el control de los actuadores, una cámara RGB modelo PI, entre otros componentes. Los robots Turtlebot3 utilizados se muestran en la Figura 8.



Figura 8: Plataforma móvil Turtlebot3® Waffle Pi.

Para implementar los esquemas de control, se utilizó la API de la plataforma ROS. ROS ofrece diversas paqueterías para acceder a los Turtlebot3®. Por medio de estos paquetes podemos

enviar las velocidades lineales v y angulares w a cada robot. Además, es posible acceder a la odometría de cada robot para medir la posición y orientación desde un marco de referencia global. También se cuenta con paquetes para medir las velocidades lineales y angulares actuales. Finalmente, los TurtleBot3® son controlados desde una PC remota por medio de ROS.

Es importante mencionar que las velocidades v y w se acotan antes de aplicarlas al TurtleBot3® para evitar dañar o desgastar los actuadores. Las acotaciones de la velocidad lineal son $-0.2 \leq v \leq 0.2$ y para la velocidad angular $-0.9 \leq w \leq 0.9$. Los resultados de la prueba se muestran a continuación.

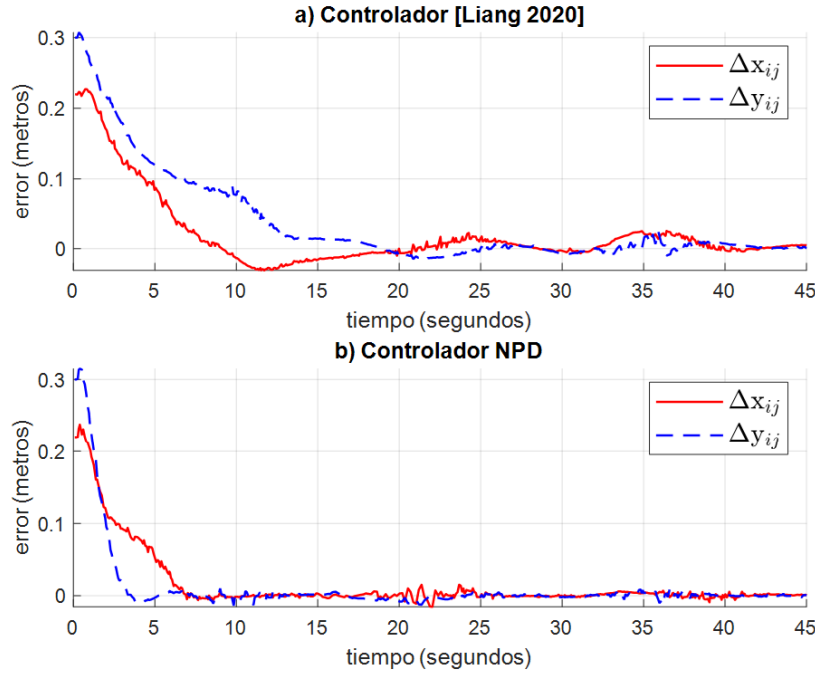


Figura 9: Errores de formación $\Delta \mathbf{r}_{ij}^*$. a) muestra los resultados de controlador clásico, mientras que b) muestra los resultados del control propuesto.

La Figura 9 muestra los errores de formación $\Delta \mathbf{r}_{ij}^*$ para las estrategias de formación comparadas. La Figura 9 (a) reporta el caso del controlador [Liang 2020], donde el error de formación converge alrededor de 20 segundos y se mantiene acotado. Pero, el error de formación es mucho mayor al error de formación de las simulaciones. Lo que sugiere que es necesario ajustar las ganancias del controlador para reducir los errores de formación. Así mismo, la Figura 9 (b) reporta el caso del controlador propuesto NPD. Como se esperaba por los resultados de las simulaciones, el error de convergencia es más rápido que el controlador clásico. En este caso, el error de formación converge a un vecindario cercano a 0 poco después de 5 segundos. Los errores de formación de simulación y experimentación son parecidos. Esto indica que la neurona adapta los pesos para compensar cualquier perturbación, ruido o dinámica no modelada, tratando de tener un desempeño similar al de simulación. Esto sugiere que no es necesario ajustar ningún parámetro a menos que se desee mejorar el desempeño del controlador NPD.

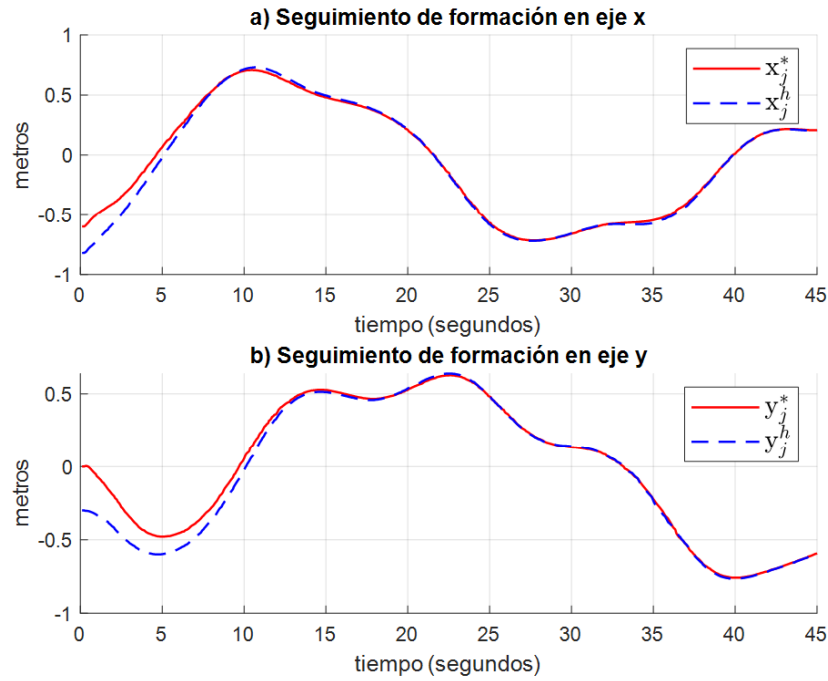


Figura 10: Seguimiento de formación del controlador clásico.

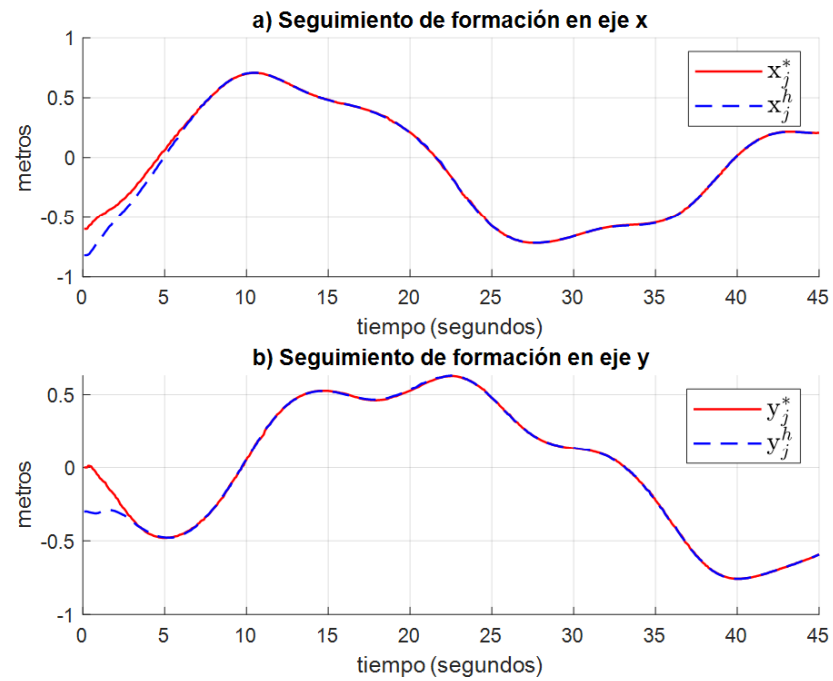


Figura 11: Seguimiento de formación del controlador propuesto NPD.

La Figura 10 muestra los resultados de seguimiento de formación del controlador [Liang 2020]. Aunque el error de formación converge después de 20 segundos, es posible notar el error de seguimiento, sobre todo en el eje x. Mientras tanto, la Figura 11 reporta los resultados de seguimiento de formación del controlador NPD. Como se esperaba, el seguimiento de formación es

muy parecido al seguimiento en la simulación. Poco después de 5 segundos, la posición actual alcanza la referencia. De acuerdo con estos resultados, el controlador propuesto NPD no solo resuelve el seguimiento en menor tiempo, sino que también demuestra un desempeño similar al de simulación, sin necesidad de un ajuste.

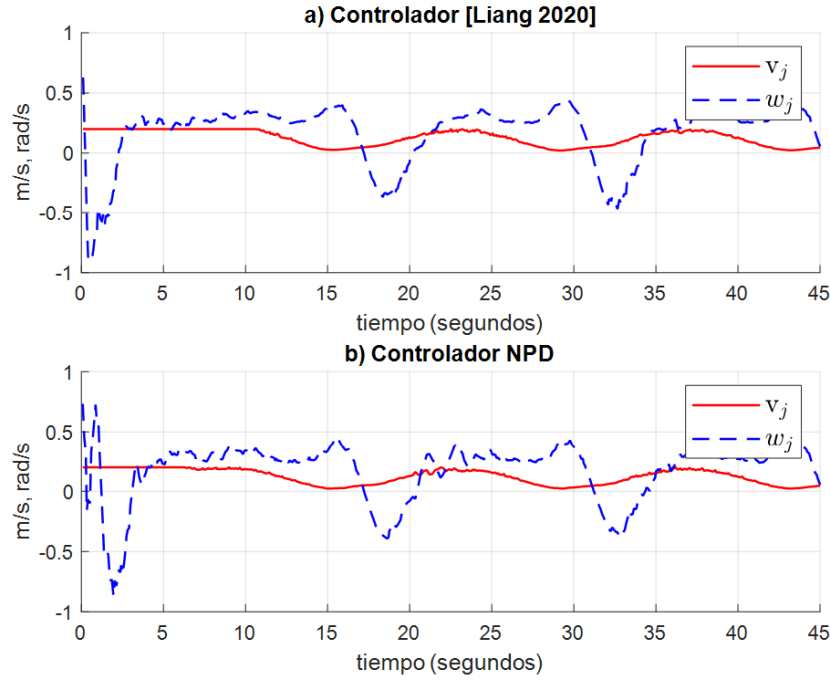


Figura 12: Acción de control para robot móvil seguidor j . a) muestra los resultados de controlador clásico, mientras que b) muestra los resultados del control propuesto. m: metros, rad: radianes, s: segundos.

Para terminar con los resultados de experimentación, la acción de control para el robot móvil seguidor se reporta en la Figura 12. Aunque la acción de control en ambos controladores tiene un patrón similar al de las simulaciones, se nota la presencia de ruido en las mediciones y perturbaciones. De acuerdo con los resultados del controlador [Liang 2020] en la Figura 12 (a), se puede observar que la acción de control v_j se mantuvo acotada hasta los primeros 10 segundos. Por otra parte, los resultados del controlador NPD en la Figura 12 (b), la acción de control v_j se mantuvo acotada solo en los primeros 6 segundos. Esto demuestra que el robot seguidor alcanzó más rápido la referencia con el controlador NPD. Los resultados del controlador NPD también muestran el efector de adaptación en la velocidad angular w_j ante la presencia de perturbaciones.

5. Discusión

De acuerdo con los resultados de la prueba de simulación, tanto el controlador [Liang 2020] como el controlador propuesto NPD reportan un seguimiento de formación adecuado. Es decir, ambos controladores llevan la posición del seguidor actual $\mathbf{r}_{ij}^h$ a la posición de referencia $\mathbf{r}_{ij}^*$ con un error de formación $\Delta \mathbf{r}_{ij}^*$ que tiende a 0. Sin embargo, el controlador NPD reporta un tiempo de convergencia menor comparado con el controlador [Liang 2020].

Por otra parte, de acuerdo con los resultados de la prueba de experimentación, el desempeño del controlador [Liang 2020] empeoró debido a que reportó un error de formación $\Delta \mathbf{r}_{ij}^*$ mayor al obtenido en la simulación. Esto indica que el controlador [Liang 2020] requiere del ajuste de ganancias para mejorar el seguimiento de formación. En contraste, el controlador propuesto NPD presentó errores de formación $\Delta \mathbf{r}_{ij}^*$ muy parecidos a los errores de formación en simulación. Esto significa que no se requiere ningún ajuste para el controlador NPD. Además, la acción de control de ambos esquemas de control muestra la presencia de ruido de medición y perturbaciones externas. Esto demuestra que el entrenamiento de la neurona ajusta sus pesos para compensar cualquier perturbación, ruido o dinámica no modelada, tratando de tener un desempeño similar al de simulación.

Las estrategias de control presentadas en este trabajo asumen que se tendrán disponibles la pose (posición y orientación) del robot móvil seguidor y el robot líder. Además, se asume que las velocidades del líder son medibles. En la práctica, no siempre es posible acceder a todas las mediciones. Por otra parte, la odometría mediante el uso de sensores abordo como codificadores ópticos o módulos IMU suelen carecer de precisión. Para mejorar la estimación de la pose de los robots móviles, es conveniente trabajar con otros esquemas como auto localización mediante el uso de cámaras y marcadores de entorno (Yu, et al., 2019). Otra estrategia útil para la estimación de la pose podría ser la técnica SLAM (Kolhatkar & Wagle, 2021). Se deja como trabajo futuro incluir algunas de estas estrategias para el cálculo de la pose y velocidades de cada robot, para el propósito de mejorar la precisión de la formación. Además de que es posible plantear una estrategia de control que no requiere del uso de las velocidades del robot líder.

6. Conclusiones

Este trabajo presentó un controlador para la formación Líder-Seguidor con base en un controlador Neuro Proporcional-Derivativo (NPD). El controlador NPD hace uso del Filtro de Kalman Extendido (FKE) para ajustar sus pesos en línea. El desempeño del controlador propuesto se validó mediante simulaciones y experimentos. Además, se incluyó una comparación con un controlador para la formación Líder-Seguidor del estado del arte. Finalmente, dos robots móviles modelo Turtlebot3® Waffle Pi se utilizaron en la prueba de experimentación mediante la plataforma ROS.

De acuerdo con los resultados reportados, el controlador propuesto NPD presentó errores de formación similares, tanto en la prueba de experimentación y la prueba de simulación. Incluso cuando la acción de control de ambos esquemas de control muestra la presencia de ruido de medición y perturbaciones externas. Esto demuestra que el entrenamiento con base en el FKE compensa las perturbaciones, ruido o dinámica no modelada. En contraste, el error de formación del controlador del estado del arte empeoró en la prueba experimental, lo que indica que este controlador requiere del ajuste de ganancias para mejorar el seguimiento de formación. Además, el controlador NPD reporta un tiempo de convergencia menor comparado con el controlador del estado del arte.

Podemos concluir que el esquema de control adaptativo NPD mejora el desempeño del controlador del estado del arte, con respecto a menor tiempo de convergencia del error de formación y capacidad de ajuste de ganancias en línea. Esto tiene la ventaja de evitar ajuste de ganancias, ante la presencia de ruido o perturbaciones externas.

Aunque este trabajo resuelve la formación Líder-Seguidor utilizando robots móviles diferenciales, como trabajo futuro resulta atractivo llevar este enfoque de formación, al uso de robots móviles omnidireccionales. Además, se pretende extender el uso del controlador NPD para resolver problemas de formación con base en consenso, utilizando cuatro o más robots Turtlebot3® Waffle Pi.

Referencias

- Alfaro, A., & Morán, A. (2020). Leader-Follower Formation Control of Nonholonomic Mobile Robots. En 2020 IEEE ANDESCON, 1-6.
- Ben-Ari, M., & Mondada, F. (2017). Elements of Robotics. Springer International Publishing.
- Bryson, J. J. (2019). The Past Decade and Future of AI's Impact on Society. BBVA. Obtenido de Open Mind BBVA.
- Dai, S.-L., He, S., Chen, X., & Jin, X. (2020). Adaptive Leader–Follower Formation Control of Nonholonomic Mobile Robots With Prescribed Transient and Steady-State Performance. IEEE Transactions on Industrial Informatics, 3662-3671.
- Hassan, M. F., & Hammuda, M. (2020). Leader-follower formation control of mobile nonholonomic robots via a new observer-based controller. International Journal of Systems Science, 1243-1265.
- Hernandez-Alvarado, R., Garcia-Valdovinos, L. G., Salgado-Jimenez, T., Gómez-Espinosa, A., & Fonseca-Navarro, F. (2016). Neural Network-Based Self-Tuning PID Control for Underwater Vehicles. Sensors.
- Hernandez-Barragan, J., Rios, J. D., Alanis, A. Y., Lopez-Franco, C., Gomez-Avila, J., & Arana-Daniel, N. (2020). Adaptive single neuron anti-windup PID controller based on the extended Kalman filter algorithm. MDPI Electronics, 636.
- Hernandez-Barragan, J., Rios, J. D., Gomez-Avila, J., Arana-Daniel, N., Lopez-Franco, C., & Alanis, A. Y. (2021). Adaptive neural PD controllers for mobile manipulator trajectory tracking. PeerJ Computer Science, 7, e393. doi:10.7717/peerj-cs.393
- Kamel, M. A., Yu, X., & Zhang, Y. (2020). Formation control and coordination of multiple unmanned ground vehicles in normal and faulty situations: A review. Annual Reviews in Control, 49, 128-144. doi:https://doi.org/10.1016/j.arcontrol.2020.02.001
- Klancar, G., Zdesar, A., Blazic, S., & Skrjanc, I. (2017). Wheeled mobile robotics: from fundamentals towards autonomous systems. Butterworth-Heinemann.
- Kolhatkar, C., & Wagle, K. (2021). Review of SLAM algorithms for indoor mobile robot with LIDAR and RGB-D camera technology. Innovations in electrical and electronic engineering, 397-409.
- Liang, X., Liu, Y.-H., Wang, H., Chen, W., Xing, K., & Liu, T. (2016). Leader-Following Formation Tracking Control of Mobile Robots Without Direct Position Measurements. IEEE Transactions on Automatic Control, 61(12), 4131-4137. doi:10.1109/TAC.2016.2547872

- Liang, X., Wang, H., Liu, Y.-H., Liu, Z., & Chen, W. (2020). Leader-Following Formation Control of Nonholonomic Mobile Robots With Velocity Observers. *IEEE/ASME Transactions on Mechatronics*, 25(4), 1747-1755. doi:10.1109/TMECH.2020.2990991
- Lu, Q., Miao, Z., Zhang, D., Ye, L. Y., Yang, S. X., & Su, C.-Y. (2019). Distributed Leader-follower Formation Control of Nonholonomic Mobile Robots. *IFAC-PapersOnLine*, 67-72.
- Oh, K.-K., Park, M.-C., & Ahn, H.-S. (2015). A survey of multi-agent formation control. *Automatica*, 53, 424-440. doi:https://doi.org/10.1016/j.automatica.2014.10.022
- Open Robotics. (2021). Why ROS? Obtenido de ROS: <https://www.ros.org>
- ROBOTIS. (2022). Robotis Turtlebot3 Overview. Obtenido de Robotis Turtlebot3: <https://emanual.robotis.com/docs/en/platform/turtlebot3/overview/>
- Sanchez, E. N., & Alanís, A. Y. (2006). *Redes neuronales: conceptos fundamentales y aplicaciones a control automático*. Cinvestav Unidad Guadalajara: Editorial Prentice Hall.
- Sanchez, E. N., Alanís, A. Y., & Loukianov, A. G. (2008). *Discrete-time high order neural control*. Warsaw: Springer.
- Sanfeliu, A., Hagita, N., & Saffiotti, A. (2008). Robotics and Autonomous Systems *Robotics and Autonomous Systems* 56 (2008) 793–797. Network robot systems, 56(10), 793–797. doi:https://doi.org/10.1016/j.robot.2008.06.007
- Siegwart, R., Nourbakhsh, I. R., & Scaramuzza, D. (2011). *Introduction to Autonomous Mobile Robots*. MIT Press.
- Tang, W., Wang, L., Gu, J., & Gu, Y. (2020). Single Neural Adaptive PID Control for Small UAV Micro-Turbojet Engine. *Sensors*.
- Wang, H., Guo, D., Liang, X., Chen, W., Hu, G., & Leang, K. K. (2016). Adaptive Vision-Based Leader-Follower Formation Control of Mobile Robots. *IEEE Transactions on Industrial Electronics*, 64(4). doi:10.1109/TIE.2016.2631514
- Xuan-Mung, N., & Hong, S. K. (2019). Robust adaptive formation control of quadcopters based on a leader–follower approach. *International Journal of Advanced Robotic Systems*, 1-11.
- Yu, H., Fu, Q., Yang, Z., Tan, L., Sun, W., & Sun, M. (2019). Robust Robot Pose Estimation for Challenging Scenes With an RGB-D Camera. *IEEE Sensors Journal*, 2217-2229.
- Zhong, J., Zhu, Y., Zhao, C., Han, Z., & Zhang, X. (2020). Position tracking of a pneumatic-muscle-driven rehabilitation robot by a single neuron tuned PID controller. *Complexity*.



Recibido 18 Nov. 2022

ReCIBE, Año 12 No.1, May. 2023

Aceptado 30 Nov. 2022

Sistema de Identificación de Incendios en Espacios Interiores

Fire Identification System in Interior Spaces

Héctor Caballero Hernández¹

Marco Antonio Ramos Corchado¹

Vianney Muñoz Jiménez¹

¹Universidad Autónoma del Estado de México

RESUMEN

Los incendios en espacios cerrados se caracterizan por ser detonados cuando alguna instalación eléctrica está provocando un cortocircuito o por una fuente que genera una chispa en dichos espacios, además, hay presencia de combustibles como gases y materiales sólidos o líquidos altamente flamables. Esta investigación se enfoca en presentar un sistema híbrido de prevención y detección de incendios en lugares cerrados, como casa-habitación, edificios corporativos, laboratorios, entre otros. Se ha propuesto la implementación de sensores para la detección de distintos tipos de gases flamables, así como la detección de gases no flamables que son producto de un proceso de combustión, aunado a esto, se cuenta con un sistema de detección por visión de computadora, el cual emplea una red neuronal profunda para la detección de fuego y humo basada en el modelo YOLO (You Only Look Once). Las variables obtenidas por el dispositivo, denominado como DRI3 (Dispositivo de reconocimiento de incendios), tiene la capacidad de almacenar los registros en un servidor local, así como subir los datos obtenidos por los sensores en la plataforma ThingSpeak para el respaldo y análisis de la información. De acuerdo con los resultados obtenidos, el sistema tuvo una capacidad de detección de gases flamables y humo en un 100%, mientras que, la red convolucional empleada para detectar fuego en imágenes digitales tuvo una capacidad de detección del 93%, adicionalmente, la emisión de alertas mediante el envío de mensajes de alerta a celulares se gestionó de forma correcta.

Palabras clave: Detección de incendios, redes neuronales convolucionales, prevención

ABSTRACT

Fires in closed spaces are characterized by being detonated when an electrical installation is causing a short circuit or by a source that causes a spark in said spaces, in addition, there is the presence of fuels such as gases and highly flammable solid or liquid materials. This research focuses on presenting a hybrid system for the prevention and detection of fires in closed places, such as a homeroom, corporate buildings, laboratories, among others. The implementation of sensors for the detection of different types of flammable gases has been proposed, as well as the detection of non-flammable gases that are the product of a combustion process, in addition to this, there is a computer vision detection system, the which uses a deep neural network for fire and smoke detection based on the You Only Look Once (YOLO) model. The variables obtained by the device, called IRD3 (Image Recognition Device), can store the records on a local server, as well as upload the data obtained to the ThingSpeak platform for information backup and analysis. According to the results obtained, the system had a 100% detection capacity of flammable gases and smoke, while the convolutional network used to detect fire in digital images had a detection capacity of 93%, additionally, the emission of alerts by sending alert messages to cell phones was managed correctly.

KEYWORDS: Fire detection, convolutional neural networks, prevention

1. INTRODUCCIÓN

El presente trabajo muestra el desarrollo de un sistema híbrido para la detección de incendios en espacios cerrados mediante reconocimiento de evento en imágenes digitales y la detección de gases y elementos que son productos de la combustión. Este documento está compuesto de las siguientes secciones, la primera sección está dedicada a mostrar un panorama general sobre la temática de incendios, la sección 2 trata sobre definiciones y conceptos básicos referentes a la temática central de sistemas clasificadores, incendios, redes de sensores, entre otros, la sección 3, describe las investigaciones que abordan el tratamiento de incendios forestales, la sección 4 se enfoca en la presentación de la metodología a seguir para construir un sistema de detección de incendios en

interiores, la sección 5 aborda el desarrollo de la propuesta general de un sistema híbrido de detección de incendios, la sección 6 presenta las pruebas y resultados que se obtuvieron para validar el sistema de detección de incendios, en las secciones 7 y 8 se desglosa el curso a seguir para la presente investigación así como las conclusiones obtenidas tras la ejecución de las pruebas y resultados, finalmente, se presentan las referencias empleadas en este trabajo.

Uno de los peligros más grandes que se presentan en lugares cerrados, como casa habitación, sitios comerciales, edificios corporativos, laboratorios, entre otros lugares, es la generación de un incendio sin control cuando existe presencia de objetos flamables en estado sólido, líquido o gaseoso y una deficiente instalación eléctrica o algún otro elemento que cause un corto circuito o temperaturas elevadas, debido a que destruyen una gran cantidad de bienes materiales, así como poner en peligro la vida de seres humanos (CONAPCI, 2022). Se estima que de cada 100 incendios que se presentan en zonas urbanas de México, el 55.6% son de casa habitación, el 34.1% en lugares de comercio o uso mixto para la vivienda, mientras que el resto se ubica en hoteles, edificios industriales y otro tipo de infraestructura. En México, aproximadamente cada año 128,000 personas sufren quemaduras de distintos grados en el cuerpo, de los casos totales el 30% son niños (Gómez, 2022), dichas lesiones condicionan su integración a la sociedad, dejando secuelas físicas y emocionales que deterioran su calidad de vida. Los casos de quemaduras graves son de aproximadamente 13,000 personas por año, de las cuales el 8% corresponde a infantes (Secretaría de Salud, 2021).

Los incendios en espacios interiores generalmente son causados por una serie de descuidos con respecto a un uso incorrecto o almacenamiento de materiales flamables, entre los que se encuentran, los gases de tipo LP, metano y butano, el uso de solventes, alcoholes, formaldehidos, resinas y otros tipos de derivados del petróleo (principalmente gasolina y diésel) (Pavez, 2021). Se tiene que una de las principales causas de los incendios es por la fuga de gas en tubería o de cilindros, los cuales no han recibido un mantenimiento rutinario de forma correcta. Según la Organización Mundial de la Salud (OMS) fallecen aproximadamente 180,000 personas a nivel mundial (OMS, 2018) por incendios.

Debido a la complejidad de los incendios en interiores existen distintas propuestas para atender la prevención y detección de estos, mediante el uso de sistemas de visión por computadora empleando redes preentrenadas (como YOLO) para detección de eventos (Park & Ko, 2020), uso de redes de dispositivos IoT (Internet of Things) en la predicción de incendios en ciudades inteligentes mediante la detección de variables como gases procedentes de combustión, entre otros elementos (Zang et al., 2021), (Lule et al., 2020), así como la combinación de visión por computadora y dispositivos IoT (Arshad et al., 2019), (Sharma et al., 2020). Siendo que los sistemas híbridos (dedicados a combinar distintas técnicas de detección de incendios), los más utilizados actualmente, por su capacidad de detectar distintas variables que indican la presencia de un incendio. En siguiente sección se presentan los elementos fundamentales de las tecnologías que se utilizan mayoritariamente para la detección de eventos en espacios interiores, por otra parte, en la sección 3 se citan trabajos actuales sobre propuestas que tratan la prevención y la detección oportuna de incendios en espacios interiores.

2. FUNDAMENTO TEÓRICO

Los sistemas de detección de eventos están basados en distintas propuestas para determinar la presencia de un fenómeno. Generalmente, los sistemas de detección de incendios se basan en el monitoreo de variables como humo y temperatura, mientras que las propuestas más modernas llegan a emplear reconocimiento visual y mediante emisión de ondas infrarrojas para clasificar la fuente calor que está generando un posible problema.

Un sensor se define como un dispositivo capaz de detectar materiales y señales para enviar una señal y permitir que se realice un proceso o continúe con este (Thomazini & de Albuquerque, 2020). Una de las aplicaciones más novedosas que existen de sensores en espacios cerrados y abiertos, es el uso de redes de sensores, las cuales son un conjunto de sensores que están interconectados mediante una pasarela para el intercambio de información a una estación de trabajo para su recolección y tratamiento (Mendoza et al., 2020), como se puede apreciar en la figura 1. Las redes de sensores pueden transmitir la información de forma alámbrica o de forma inalámbrica (RIS).

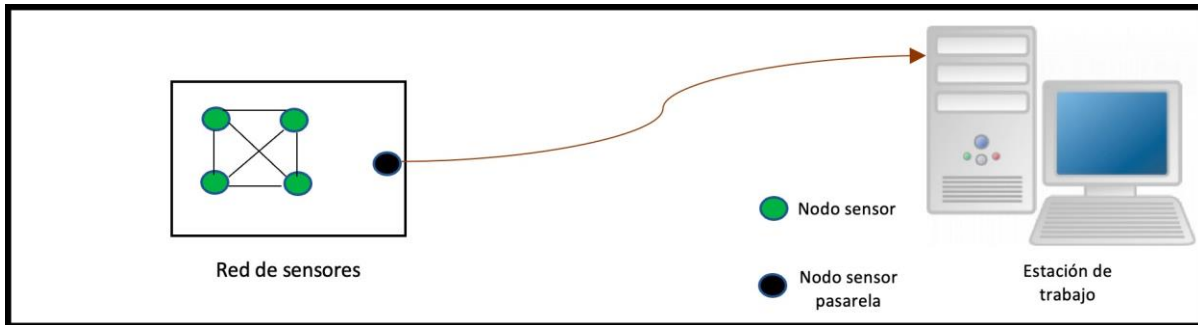


Figura 1. Esquema general de una red de sensores
Fuente: (Campaña Bastidas, 2019)

Por otra parte, en el reconocimiento de objetos y eventos, se tiene el uso de redes neuronales, específicamente, el uso de redes neuronales convolucionales (en inglés Convolutional Neural Network, CNN), las cuales son algoritmos que están basados en el reconocimiento de patrones específicos, mediante la formación de un kernel que indica una serie de características que se pueden presentar en una imagen digital (Cetinic, Lipic & Grgic, 2020).

Existen distintos tipos de CNN preentrenadas, entre las más comunes se tiene a R-CNN (Girshick et al., 2013), YOLO (Redmon, 2016), SSD (Liu et al. 2015) y VGG (Simonyan & Zisserman, 2014). En la figura 2 se encuentra la estructura básica de una red neuronal convolucional, en donde los datos de entrada pasan por un proceso de convolución, posteriormente los datos son expuestos a un proceso de mezclado y finalmente se realizan las operaciones de procesamiento en la capa de clasificación para obtener una salida.

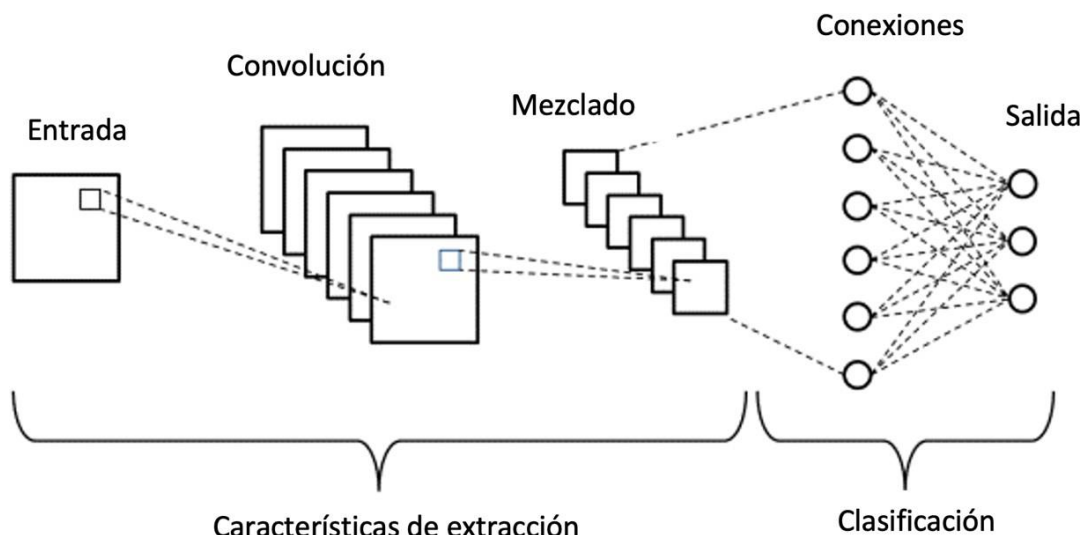


Figura 2. Estructura de una red neuronal convolucional
Fuente: Elaboración propia

Algunas de las métricas de evaluación más importantes que se utilizan en el entrenamiento de redes neuronales son las siguientes (Sze, 2020).

Accuracy. Indica la proporción de aciertos que ha tenido el modelo. Es el porcentaje de TP y TN respecto al total.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \quad (1)$$

Precision. Indica el porcentaje de los elementos válidos respecto a los predichos como verdaderos.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall. Indica la capacidad del modelo para identificar un suceso.

$$\text{Recall} = 2 * \frac{TP}{TP + FN} \quad (3)$$

F1-score. Es la media armónica entre Recall y Precision. Se emplea cuando se necesita saber la relación entre los errores de falsos positivos y falsos negativos.

$$F1 - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Donde:

True Positives (TP)
True Negatives (TN)
False Positives (FP)
False Negatives (FN)

En la siguiente sección se presentan trabajos relevantes relacionados con la detección de incendios en espacios interiores.

3. ESTADO DEL ARTE

Con los avances obtenidos en ciencias computacionales y la electrónica, ha sido posible la implementación de avanzados modelos matemáticos que permitan la detección y predicción de eventos, como es el caso de incendios en lugares donde habitan personas para evitar mayores percances. En el trabajo de Rzaïj y Al-Obaidi (2022) desarrollaron un análisis de 12 parámetros para verificar el riesgo que existe de presentar en edificios, algunos de los parámetros contemplados son, tipo de construcción, alarmas de incendios, salidas de emergencias, entre otros. Una propuesta interesante para el control de incendios está expuesta en Wang et al. (2022), donde emplean imágenes digitales para realizar una predicción de incendios, tomando en cuenta la evolución del humo que se está detectando, y en una evaluación de más de 1800 escenarios de incendios simulados, las imágenes son analizadas por una red neuronal convolucional. El factor clave que los autores analizan es el HRR (potencia del fuego), debido a que establece una relación entre la cantidad de fuego presente y el humo que se desprende en tiempo real, en la investigación los autores obtuvieron una tasa de error menor al 20% en su fase experimental. En Kou et al. (2021) presentan un modelo para detectar y predecir la intensidad del fuego en edificios con la implementación de un GRU, con la aplicación de redes neuronales que permiten identificar elementos de un incendio, la base de entrenamiento está construida por una base de imágenes de incendios. El modelo presentado tiene una precisión de localización del fuego en un 0.954 (de 1), aunque tiene limitaciones cuando el fuego es constante y el escenario simple.

Un trabajo interesante en la evaluación de riesgos de incendios basado en aprendizaje automático y estadístico es el trabajo de Choi & Jun (2020), los autores comparan su modelo contra el KFRI, en los resultados experimentales obtuvieron una precisión del 0.7514, lo cual representa un gran esfuerzo debido a que la periodicidad con que se presentan los incendios en interiores, así como los factores que intervienen en ellos son altamente difíciles de estudiar.

En las investigaciones de Iqbal et al. (2020), Mwedzi et al. (2019) y Muhammad et al. (2018) presentan propuestas similares basadas en redes neuronales convolucionales para la detección de incendios en interiores de edificios, siendo el más avanzado el trabajo Muhammad et al. (2018), debido a que

desarrolló un framework de vigilancia contra incendios para edificios y casa habitación, incendios forestales e incendios en vehículos, la red convolucional propuesta está basada en GoogleNet, consta de 100 capas con dos redes convolucionales principales y 4 pools, la base de entrenamiento está en imágenes digitales de una profundidad de 24 bits de 224 x 224 píxeles en 64 núcleos, la novedad contra otras propuestas es que está aplicado en vídeo y tiene una efectividad de detección promedio del 94.43% en el mejor de los casos y un 80% en escenarios de alta complejidad.

Algunas investigaciones se han concentrado en el análisis de imágenes digitales desde el punto de vista de su intensidad calorífica, usando tecnología infrarroja, tal es el caso de las investigaciones de Rostami et al. (2022), Wang et al. (2018), Yang et al. (2018), Yang et al. (2020), Saponara et al. (2021) entre lo que destaca el uso de modelos avanzados de reconocimiento de imágenes digitales con imágenes normales de tipo RGB (Red Green Blue) y de imágenes digitales tomadas con cámaras infrarrojas. En otros trabajos como en Nazir et al. (2022), realizan una experimentación detallada sobre los componentes que existen espacios cerrados y analizan factores como la temperatura, la concentración de gases como CO₂, para medir los niveles de alerta que se pueden presentar.

En los trabajos de Meidelfi et al. (2022), Albassam y Khan (2019) y Kothandapani et al. (2021) emplearon sensores como MQ6 y MQ5 para ser utilizados en dispositivos móviles con Android y detectar humo. En trabajos como Raj et al. (2021) y Guo et al. (2019), en este último trabajo desarrollaron un sistema el cual consta de dos partes: terminal sensor y servidor central, y tiene un módulo receptor del sistema de posición global (GPS), un módulo de servicio general de paquetes de radio (GPRS) y un módulo de alimentación. Las redes de sensores inalámbricos móviles consisten en terminales de sensores móviles y terminales de sensores estacionarios, permiten un monitoreo de fugas.

Los sistemas de detección de fuego hacen uso de distintas herramientas para poder identificar elementos, tales como, fuentes posibles de materiales que puedan ocasionar un incendio, como lo son: gases como el butano, propano, metano, entre otros. Varias de las propuestas modernas en la identificación de incendios están basadas en el uso de análisis visual mediante CNN, así como modelos predictivos que indican el comportamiento destructivo del fuego en interiores si se llegase a presentar, algunas otras propuestas se centran en emplear redes de sensores para detectar la presencia de elementos que indicarían la presencia de humo y gas. Debido a la complejidad que representa un incendio en espacios interiores es necesaria la combinación distintas tecnologías que permitan precisar con certeza este tipo de eventos.

En base a la revisión de los trabajos anteriormente descritos, se ha diseñado una propuesta para la detección de incendios en espacios interiores la cual se describe a continuación.

4. METODOLOGÍA

En la detección de incendios es importante identificar los elementos que presentan cambios en el ambiente, entre las variables más importante se tienen la temperatura y la humedad, mientras que, la detección de presencia de elementos como gases inflamables, son elementales para prevenir un incidente, por otro lado, la detección de elementos como el humo o flama indicarían que existe un proceso de combustión, el cual debe de detenerse antes de que se expanda sin control. Tomando en consideración los elementos anteriores, se ha propuesto la metodología expresada en la figura 3, en la cual se indican 4 fases para la identificación de incendios o fuentes potenciales que pueden provocar un incendio en interiores.

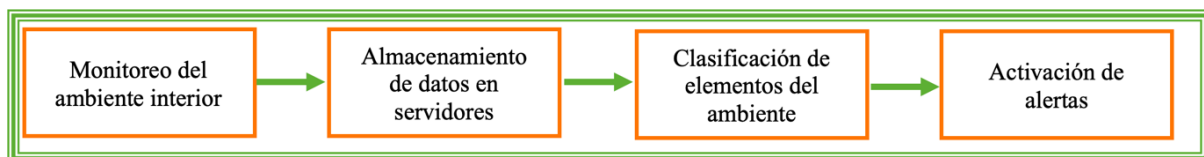


Figura 3. Metodología propuesta para la detección de incendios en interiores
Fuente: Elaboración propia

Las fases de la metodología propuesta se explican en los siguientes párrafos.

- Monitoreo del ambiente interior. Se detectan los valores de las variables como temperatura y humedad, así como la verificación de presencia de gases flamables, humo y flama. Es la fase de mayor importancia, debido a que permite obtener datos que tienen como función principal evitar la generación de un incendio o la expansión de este.

- Almacenamiento de datos en servidores. Los datos obtenidos en la fase anterior son almacenados en dispositivos locales o en la nube, para realizar un análisis a corto y mediano plazo para la caracterización del microclima y poder realizar ajustes de temperatura y humedad.

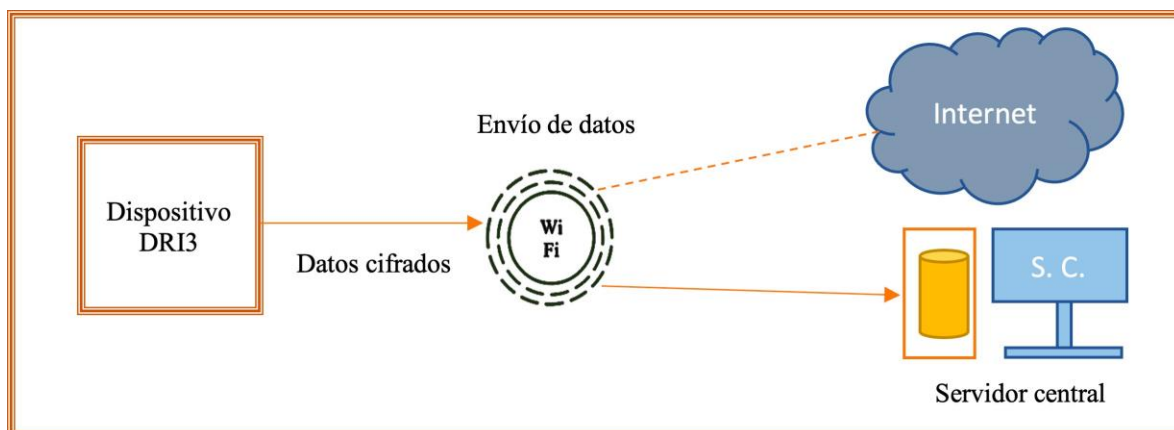
- Clasificación de los elementos del ambiente. El valor de los elementos al ser obtenidos y almacenados, son comparados con umbrales para determinar si se exceden los valores que se considerarían como estables, tanto el valor de la temperatura ambiente y la humedad, siendo prioritarios la presencia de gas flamable, humo y presencia de flama para realizar un análisis mediante visión por computadora.

- Activación de alertas. En la cuarta fase se envían alertas cuando los sistemas de detección obtuvieron presencia de gas, humo o fuego, como acción de alerta y acción para los usuarios vinculados con el espacio interior.

Las fases de la metodología desarrollada son aplicadas para la construcción de un sistema de detección de incendios, con el propósito de generar una propuesta que se adapte a distintos contextos, permitiendo flexibilidad y adaptabilidad a distintos entornos de interiores. En la sección 5 se presenta una propuesta general sobre la metodología planteada.

5. PROPUESTA GENERAL

En base a la metodología anterior se ha diseñado una propuesta que englobe la detección de variables físicas mediante sensores que registren humedad, temperatura ambiente, presencia de humo y fuego, así como un análisis de imágenes digitales. En la figura 4, se muestra el diseño general de un sistema híbrido denominado como DRI3 (Dispositivo de Reconocimiento de Incendios), el cual tiene la capacidad de obtener la lectura de las variables anteriormente mencionadas, y cuenta con la capacidad de captura imágenes para comprobaciones visuales de la presencia de fuego. Se ha propuesto que el DRI3 registre todos los elementos que son obtenidos por los sensores en una base de datos local, así como una conexión a internet, mediante una conexión a una red local, cualquier evento que se supere los umbrales predeterminados (límites a los que pueden llegar las variables físicas), iniciarían un proceso de alarma acústica, así como la activación de un aviso mediante la emisión de alertas acústicas y vía mensajes SMS.



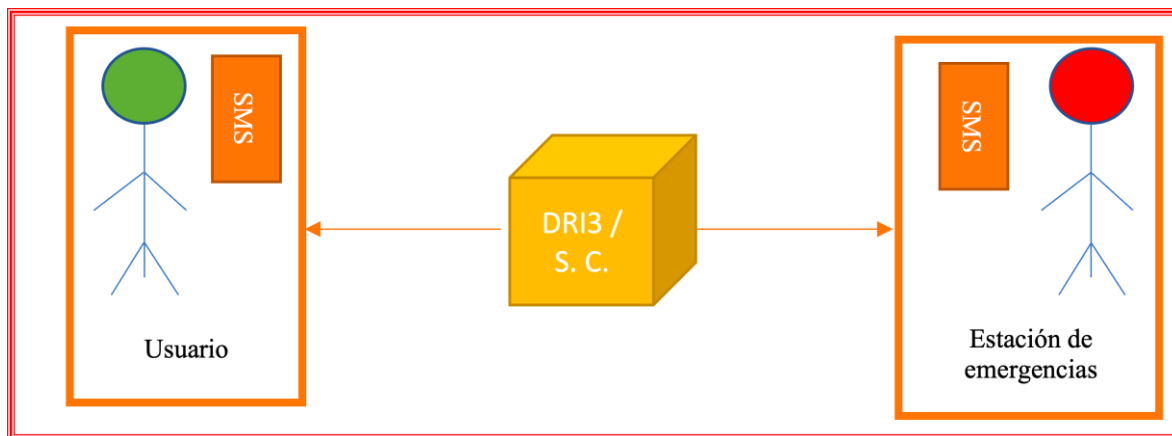


Figura 4. Propuesta general de prevención y atención a incendios en espacios interiores
Fuente: Elaboración propia

Los datos que se transmiten en la red local, son codificarlos mediante el uso de la base 64. Cuando los datos llegan al servidor local son decodificados y almacenados en la base de datos, posteriormente, se envían a la plataforma en la nube para realizar proceso de graficación de la información que se está almacenando. La información registrada por los sensores puede ser directamente cargada a la plataforma que se encuentra en la nube, pero si la conexión es inestable se perderían los datos, además depende de las circunstancias que se presenten en el entorno, por esta razón está diseñado para almacenar de forma prioritaria en un base de datos local.

En la figura 5 se muestra la estructura de bloques del DRI3 indicando los componentes con los que está conformado, el cual contiene una configuración donde el dispositivo IoT realiza la lectura de los sensores (monóxido de carbono, gas flamable, temperatura, humedad y humo), mientras que el analizador imágenes dispone de otro hardware, como puede ser un Tablet PC o PC, junto con una cámara para la captura de imágenes y un módulo de comunicación Wifi para la transmisión de información. Los dispositivos que concentran el procesamiento están conectados por un gateway, el cual es el mecanismo para la transmisión de la información, esta gateway puede ser mediante un medio inalámbrico o alámbrico para lograr la comunicación entre el dispositivo IoT y la computadora.

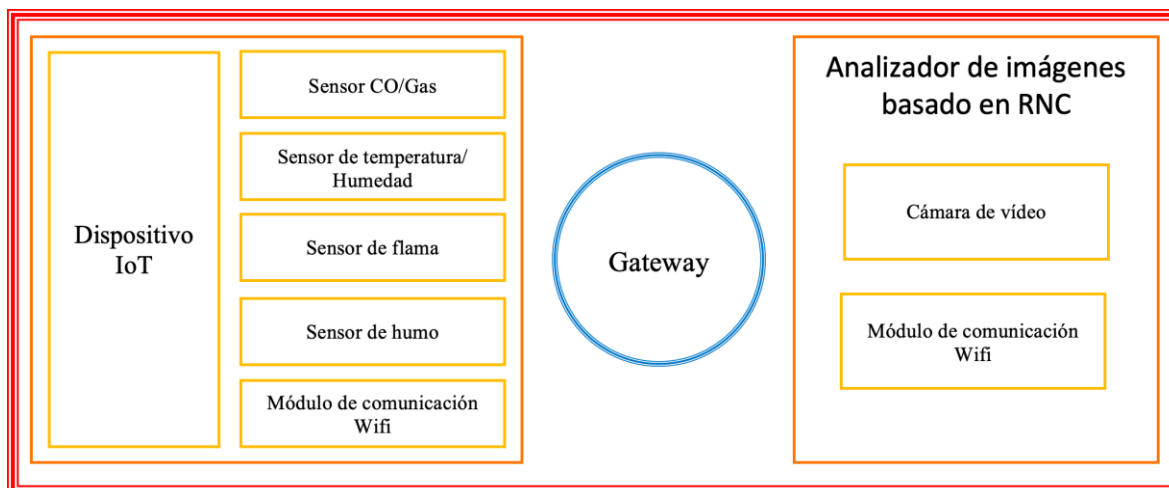


Figura 5. Componentes del dispositivo de reconocimiento de incendios para espacios interiores (DRI3)

Fuente: Elaboración propia

Los umbrales son determinados con respecto las características técnicas de los sensores, pero la lógica de activación de las alertas es la siguiente.

- 1.- Los sensores de gas y humo tienen prioridad al superar su umbral, activan directamente las alarmas disponibles y el análisis visual.
- 2.- Si cualquiera de los sensores de temperatura o flama superan su umbral, se activa el análisis visual.
- 3.- Si el análisis visual no detecta presencia de fuego, genera una alerta de inspección.
- 4.- Todas las lecturas de los sensores son almacenadas en una base datos sin importar que no se rebase el umbral.

La propuesta del DRI3 está diseñada cubrir distintas áreas de espacios cerrados, debido a que el concepto de comunicación IP permite conectar varios dispositivos de acuerdo con la capacidad de administración de conexión que provee punto de acceso. En la figura 6 se muestra una distribución hipotética del DRI3, en donde cada uno de los dispositivos tendría asignada una IP única, sirviendo para indicar el área en donde se está desarrollando un determinado evento.

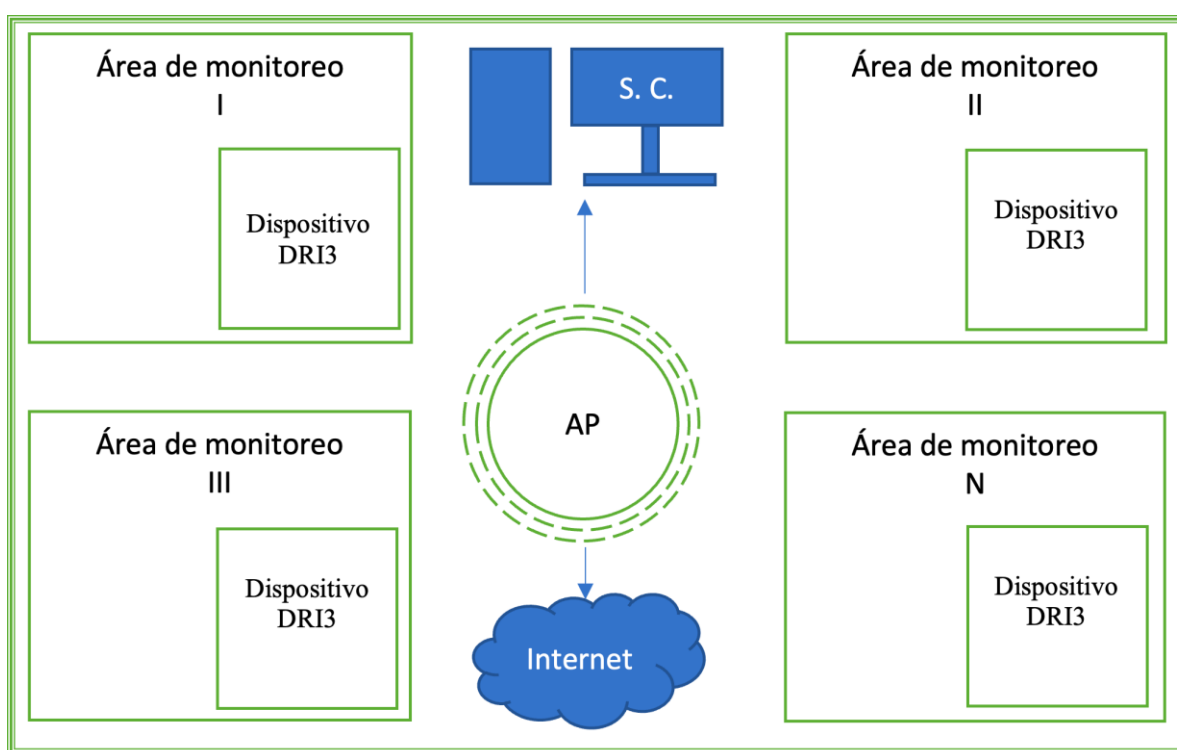


Figura 6. Distribución de DRI3 en distintos espacios
Fuente: Elaboración propia

La configuración del DRI3 depende de las necesidades específicas del ambiente en el que estaría trabajando, por esta razón se ha diseñado de la forma más flexible posible para trabajar en entornos con o sin conexión a internet, así como la cantidad de sensores que puede incluir.

6. PRUEBAS Y RESULTADOS

Las pruebas se desarrollaron en espacios interiores con dos escenarios de evaluación, ambos tomando en cuenta una casa habitación. En el primer escenario, denominado A, se libera gas LP de forma controlada para medir la capacidad de detección del DRI3. En el segundo escenario, denominado B, se produce un incendio controlado, en donde se busca detectar la presencia de humo y la activación del sistema de visión por computadora para determinar la capacidad de reconocimiento visual, teniendo en cuenta que el objeto a detectar es fuego. Los datos que se obtienen por el sistema son almacenados en las bases de datos local y en la nube, teniendo prioridad la primera. En ambos escenarios se

empieza con el proceso de detección de los datos del ambiente los cuales se tomarán como parámetros normales de operación.

La configuración del DRI3 está conformada por una placa IoT Wemos D1, la cual cuenta con una tarjeta inalámbrica ESP8266 integrada, que trabaja a una frecuencia de 2.4 GHz, a la tarjeta controladora se han conectado los sensores HI 01 (detección de flama), MQ-9 (detector de humo y gas) y DHT11 (sensor de temperatura ambiental y humedad). Por otra parte, el analizador de imágenes se montó en una tableta Chuwi Hi10 Go, la cual cuenta con un procesador Intel Celeron N5100, 6 GB de memoria RAM, con 128 GB de almacenamiento en disco SSD y doble cámara (frontal y trasera). La elección de la tableta electrónica se debe a su tamaño compacto y además que incluye cámara integrada, pero el sistema puede trabajar con cualquier otro tipo de hardware que permita capturar vídeo e imagen. La transmisión de datos entre la placa Wemos D1 y la tableta se ha realizado vía Wifi mediante un dispositivo de doble banda que trabaja a 2.4 GHz y a 5 GHz. La configuración de la gateway está dada por la conexión inalámbrica que provee el punto de acceso. Los sensores han pasado por un proceso de depuración, el cual ha consistido en la conexión de dichos dispositivos durante un periodo de 12 horas ininterrumpidas, puesto que, al momento de realizar el proceso de lectura sin una depuración previa los datos que se registran son erróneos.

La figura 7 muestra la distribución de los componentes del DRI3 de forma esquemática, en la parte de la izquierda se ubican los sensores, la tarjeta de adquisición de datos y la unidad de alimentación, mientras que en la segunda parte se presenta el dispositivo de empleado para el análisis visual. Las dos secciones del DRI3 no tienen conexión física, debido a que la comunicación se realiza de forma inalámbrica, siendo el flujo de información de forma unidireccional, el cual parte de la placa de adquisición de datos a la tableta electrónica.

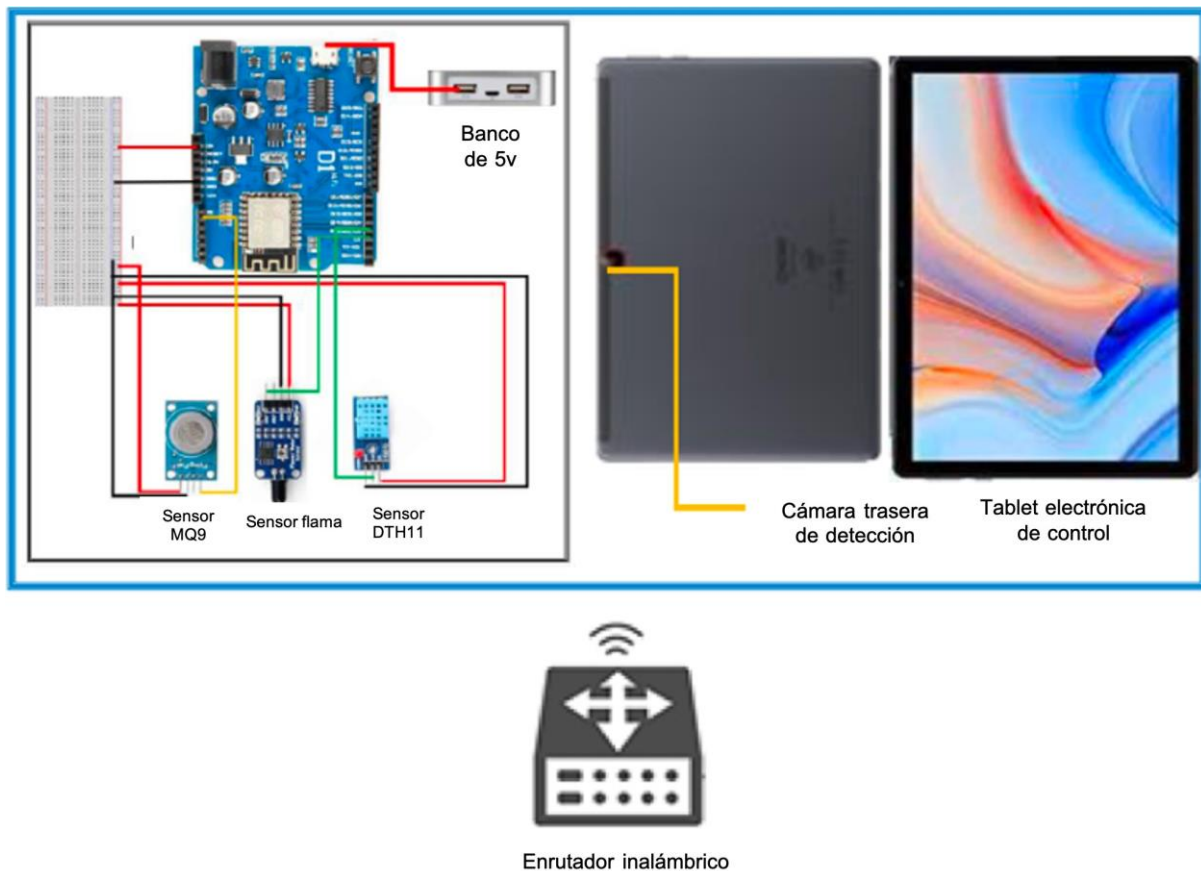


Figura 7. Esquema de los componentes del DRI3
Fuente: Elaboración propia

El modelo de reconocimiento de imágenes para detección de fuego se basó en la red convolucional YOLO v5, específicamente utilizando el modelo preentrenado yolov5l6. La base de datos utilizada para el entrenamiento de la red se construyó con imágenes digitales provenientes de internet, tomando 7000 imágenes del tipo RGB con una resolución 640 x 640 píxeles. El entrenamiento se llevó a cabo en el entorno de Google Colab, mediante una GPU Nvidia TESLA de 16 GB de RAM, empleando 100 ciclos de entrenamiento (épocas). Los resultados obtenidos del proceso de entrenamiento, que se muestran en la figura 8, presentan resultados de las métricas mPA50 y mPA95, con puntajes de 0.8 y 0.6 respectivamente. Mientras que, la métrica precision obtuvo un puntaje de 0.67 y recall obtuvo un puntaje de 0.82, indicando que los objetos detectados van a tener más de un 80% de precisión al momento de que el software realice el etiquetado de área de los objetos reconocidos.

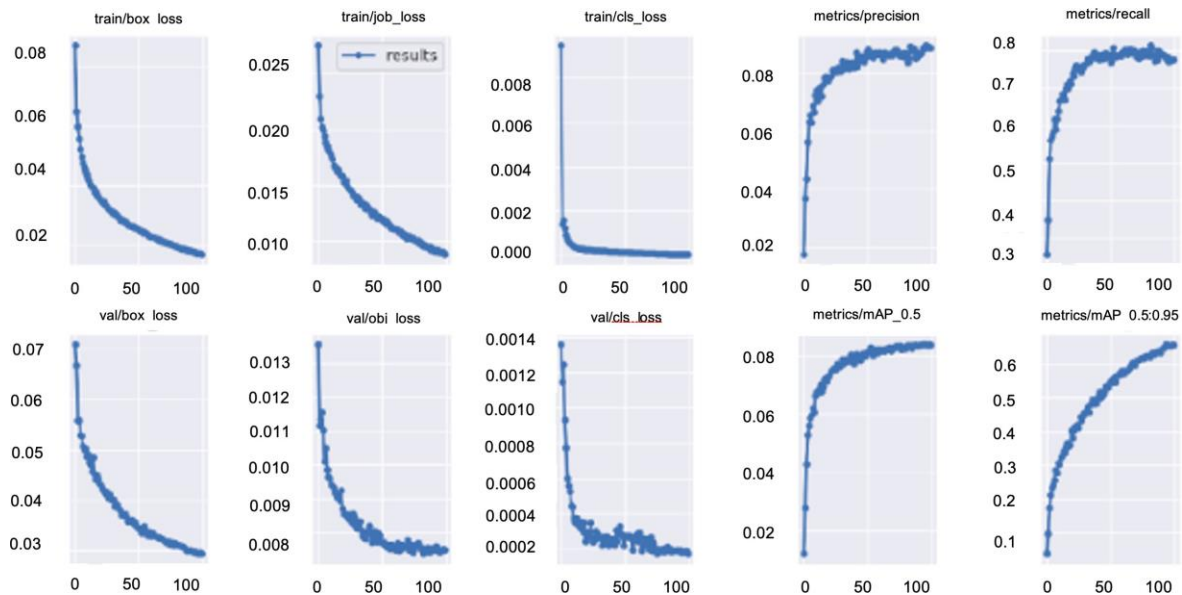


Figura 8. Resultados obtenidos del entrenamiento del modelo de CNN de detección de fuego del modelo preentrenado yolov5l6
Fuente: Elaboración propia

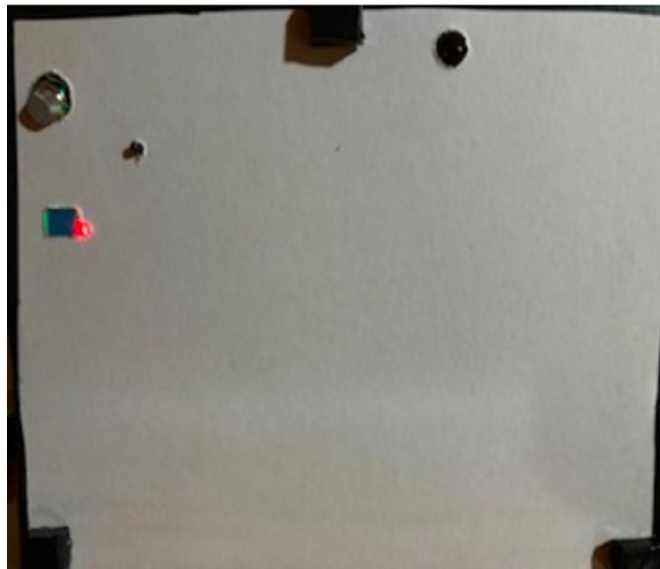
Una vez obtenido el modelo de reconocimiento de fuego, se sometió a un proceso de evaluación empleando imágenes del dataset Kaggle (Gamaleldin et al. 2018), con un total de 750 imágenes digitales con presencia explícita de fuego, para validar la efectividad del modelo, del total de imágenes se pudo detectar con éxito el 93% de las imágenes digitales, por otra parte se evaluaron 240 imágenes digitales en las cuales no existía presencia de fuego, de las cuales el 100% fueron clasificadas sin la presencia de la clase buscada.

Por otra parte, la placa de adquisición WEMOS D1 con los sensores conectados fueron puestos a prueba a 100 ciclos, para detectar la presencia de gas, humo y emisión de fuego, sin que se presentara un falso positivo o un falso negativo.

La figura 9 muestra la vista física del dispositivo DRI3 que se ha utilizado en el proceso de pruebas, como se puede observar en la imagen contenida en la figura 8 inciso a, en la parte inferior se ubican los sensores junto con la placa Wemos D1, en la sección de la derecha superior se ubica la tableta que realiza el procesamiento de visión por computadora, finalmente, en la sección de la izquierda superior se ubica el espacio para colocar las baterías que alimenten a los sensores. En las pruebas experimentales se utilizó una conexión directa a un tomacorriente, no se utilizó una batería durante estas pruebas, es por ello por lo que se observa el espacio vacío en la sección de baterías. En la figura 8 inciso b se expone la cara inferior del prototipo, en donde se aprecian los sensores y la cámara digital de la tableta electrónica, finalmente en la figura 8 inciso c se muestra una vista del prototipo en una perspectiva frontal y superior.



a) Vista del superior del prototipo en funcionamiento



b) Cara inferior del prototipo



c) Vista frontal y superior del prototipo

Figura 9. Diferentes vistas del prototipo para detección de incendios en interiores
Fuente: Elaboración propia

6.2 Escenario A

En la ejecución de la prueba del escenario A se ha podido obtener correctamente la lectura de gas LP, mediante la simulación de una fuga de este tipo de gas. En la figura 10 se presenta el mensaje que ha recibido el usuario al momento de que se detectó un evento de fuga de gas, para que pueda tomar una acción inmediata. Dentro de este proceso se ha activado la alarma acústica para alertar a las personas que se encuentren en el área afectada. El sistema de detección de incendios, al momento que se ha detectado la fuga de gas, almacena los datos en la tableta electrónica en una base datos construida en SQLite y posteriormente activa el servicio de SMS de la plataforma Twilio.

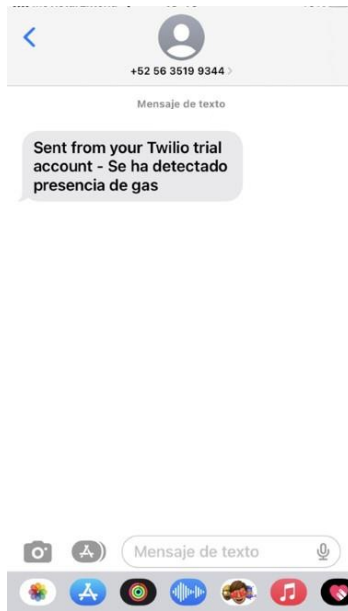


Figura 10. Mensaje recibido de la plataforma Twilio en un evento de detección de gas
Fuente: Elaboración propia

6.3 Escenario B

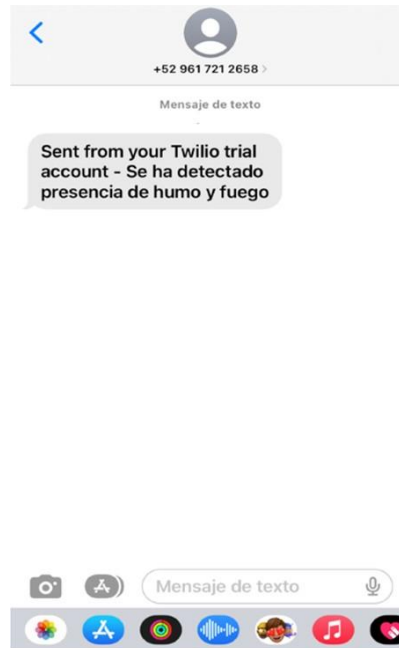
En el segundo escenario se activó el sistema de detección visual al momento de que se presentó la fase de incendio, el software fue capaz de detectar correctamente cuando se estaba produciendo el proceso de combustión. Se ha podido observar que el tiempo de reacción del sistema de reconocimiento visual oscila entre unos 10 y 15 segundos después de que el módulo de detección de gases se ha activado y por lo tanto ha emitido la alerta de activación de la comprobación visual. En la figura 11, se presentan los ejemplos de las imágenes que se capturaron al momento de la activación del sistema de detección de incendios, en los incisos a y b de la figura 11, se exponen las capturas de las tomas de los objetos en donde se detectó la presencia de fuego, enmarcando las secciones en donde se encuentran los objetos que se están incendiando. La activación del sistema de reconocimiento fue mediante la presencia de humo que detectó el sensor MQ9. El sistema de envío de datos para las alertas al usuario con mensajes SMS ha funcionado de manera correcta (plataforma Twilio), como se puede observar en la figura 11 c).



a) Detección de fuego 1



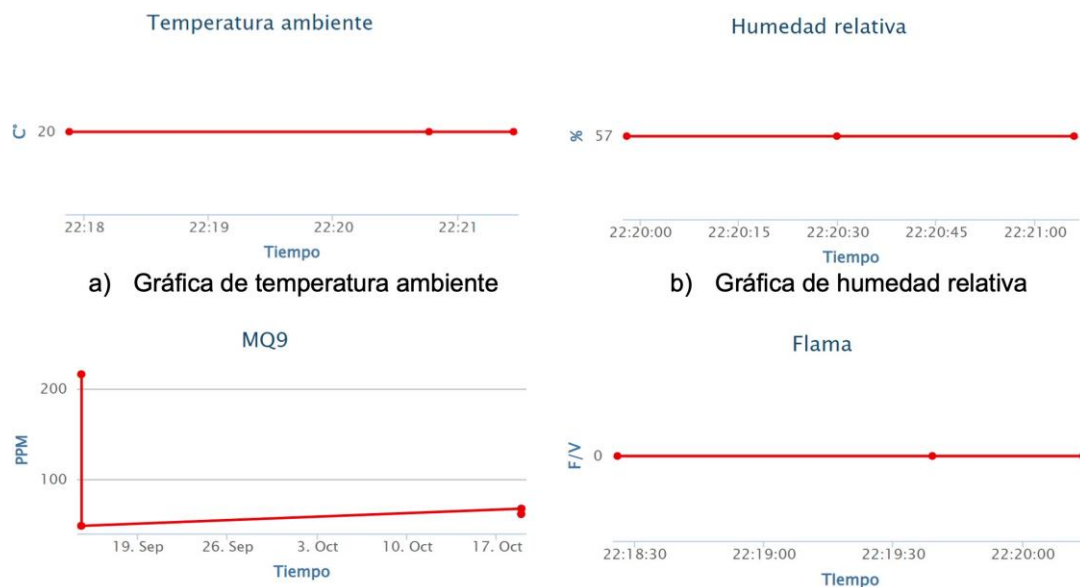
b) Detección de fuego 2



c) Mensaje de alerta de presencia de humo y fuego

Figura 11. Resultados obtenidos por el sistema de detección visual
Fuente: Elaboración propia

Los datos adquiridos por los sensores conectados a la placa WEMOS D1 fueron almacenados en la base de datos contenida en la tableta electrónica, como respaldo local, para posteriormente ser enviados al servicio en la nube ThingSpeak (de MathWorks) mediante un script de Python. Las variables almacenadas en el servicio en la nube son temperatura ambiente (figura 12 a), humedad relativa (figura 12 b), PPM de gas y humo (figura 12 c) y evento de flama (figura 12 d), para poder tener el registro de las condiciones normales del espacio en el DRI3 se encuentra trabajando.



c) Gráfica de lectura de PPM gas y humo d) Gráfica de lectura de evento de flama

Figura 12. Graficas de datos almacenados en la plataforma ThingSpeak
Fuente: Elaboración propia

De acuerdo con los resultados obtenidos tras la ejecución de las pruebas del DRI3, se puede observar que la propuesta es altamente efectiva para reconocer distintos factores que indican un suceso en una etapa temprana de incendios, así como en una etapa avanzada, puesto que, al tratarse de un sistema híbrido que emplea las ventajas del reconocimiento de fuego mediante visión por computadora, así como el uso de sensores electroquímicos, permite analizar distintas perspectivas al momento de indicar que se está presentando un evento de incendio, con una respuesta efectiva menor a un 1 minuto, desde el momento que los datos obtenidos son verificados, hasta que el usuario final recibe la alerta vía SMS y sonora.

La propuesta desarrollada tiene la ventaja de trabajar con la arquitectura del protocolo IP, lo cual le permite desplegar, de manera teórica, una gran cantidad de dispositivos en interiores para cubrir distintos puntos estratégicos, así como el uso de distintos tipos de hardware para realizar la lectura de gases, temperatura, humedad y flama, aprovechando la arquitectura de hardware abierto que proveen plataformas como Arduino y sus derivados, por otra parte, el software empleado para el reconocimiento visual puede implementarse en distintos tipos de hardware que corran con los sistemas operativos Windows o Linux, siendo que DRI3 está concebido para adaptarse a los entornos donde debe desempeñarse, sin estar limitado a un tipo específico de hardware y de tipo de sistema operativo. Una de las cualidades de DRI3 sobre propuestas del estado del arte, anteriormente analizadas, es que, al capturar los datos del ambiente de los sitios en donde se encuentra desplegado, permite evitar errores por fluctuaciones de elementos como la temperatura y la humedad, debido a que no permanecen constantes en todos los edificios, así como el clima en donde se encuentra puede variar, sin que ello indique que se esté presentando un incendio.

Los sistemas híbridos de detección de fuego tienen la ventaja de obtener información desde distintas perspectivas, debido a que en este caso particular, se realiza una detección empleando visión por computadora para analizar las formas del fuego, y por otra parte, el uso de sensores electroquímicos permite detectar variables como humo, fuego y temperaturas elevadas, para dar una respuesta complementaria al análisis de visión por computadora, sumando que no hay restricciones con un respecto al uso de hardware específico y la adecuación previa al clima en donde se encuentra el DRI3.

7. TRABAJO A FUTURO

De acuerdo con los resultados obtenidos del funcionamiento del DRI3, es posible ver que la detección de elementos como gas y humo, así como la detección de fuego vía imágenes digitales, ha cumplido satisfactoriamente su objetivo primario, pero existen limitaciones con respecto a la distancia de transmisión del dispositivo ESP8206, debido a que en ambientes con objetos densos la transmisión de datos oscila entre 15 y 25 metros, mientras que en espacios donde no existen obstáculos para las señales inalámbricas puede superar los 30 metros de transmisión, por lo tanto, se propone analizar la adquisición de otros dispositivos de transmisión de datos que incrementen la distancia de comunicación. El software de análisis visual trabajó de forma exitosa, pero se ha considerado incluir la clase que contemple la detección de humo, para analizar situaciones en las que no exista visiblemente presencia de fuego.

Por otra parte, se considera la conexión del DRI3 a dispositivos de control de incendios, como pueden ser aspersores de agua y de espuma antincendios, así como la activación de extractores de gases para eliminar la concentración de gases inflamables en interiores.

Finalmente, en investigaciones futuras se tiene contemplado realizar el despliegue de varios DRI3 en diversas zonas para verificar su eficacia práctica, en la presente investigación no fue posible realizarlo por las limitaciones presupuestales con las que se contaban.

8. CONCLUSIONES

Las propuestas dedicadas a la prevención y detección de incendios en espacios cerrados han tenido un gran avance, debido a que las modernas implementaciones en redes neuronales convolucionales permiten el reconocimiento de factores como el fuego, humo, así como identificación de zonas de

riesgo. Por otra parte, las redes de sensores han tenido una evolución lo suficientemente razonable para poder incorporar una gran cantidad de elementos que ayuden a detectar el cambio de variables en el entorno que indiquen un posible incendio o la presencia de uno.

De acuerdo con los resultados obtenidos, es posible visualizar que la propuesta inicial resultó efectiva para la detección de fuego en su etapa temprana, así como la detección de formas como son llamas, el modelo basado en la red convolucional YOLOv5, durante la fase de validación del modelo obtuvo una precisión del 93 % con respecto a las imágenes que pudo clasificar correctamente del dataset Kaggle.

Durante la ejecución del módulo de visión artificial se observó que el uso constante de este genera un alto costo de consumo energético y de recursos computacionales (tiempo en CPU), debido a que suele demandar grandes cantidades de tiempo en procesar la información adquirida.

Uno de los grandes retos a los que se enfrentan este tipo de propuestas es el costo de los dispositivos para realizar el procesamiento adecuado, debido a que el análisis por computadora de imágenes digitales demanda altas tasas de consumo en procesador, y, por otra parte, los sensores deben de utilizarse en puntos estratégicos que permitan monitorear gases como el CO, CO2 y los niveles de temperatura cuando se elevan de forma drástica.

El modelo que emplea el DRI3 le permite combinar hardware de distintas características, para adaptarse a situaciones del entorno en el cual se va a desplegar, así como la consideración de condiciones ambientales, las cuales son almacenadas de forma local o mediante un servicio en la nube para tomar los parámetros de referencia, esta última característica generalmente no es tomada en cuenta en las propuestas del estado del arte.

Agradecimientos

Agradecemos al Consejo Mexiquense de Ciencia y Tecnología por el apoyo otorgado para la realización del de este proyecto de investigación.

REFERENCIAS

- Albassam, N., & Khan, J. (2019). Autonomous fire safety system for gas leak detection. doi:10.47611/JSR.VI.898.
- Arshad, B., Ogie, R., Barthelemy, J., Pradhan, B., Verstaavel, N., & Perez, P. (2019). Computer vision and IoT-based sensors in flood monitoring and mapping: A systematic review. *Sensors*, 19(22), 5012.
- Campaña Bastidas, S. E., Cabrera Meza, H. E., Cervelion Bastidas, A. J., & Aguirre Cabrera, A. (2019). Capítulo 1: Las redes de sensores inalámbricas, arquitectura y aplicaciones.
- Cetinic, E., Lipic, T., & Grgic, S. (2020). Learning the principles of art history with convolutional neural networks. *Pattern Recognition Letters*, 129, 56-62.
- Choi, Myoung-Young & Jun, Sunghae. (2020). Fire Risk Assessment Models Using Statistical Machine Learning and Optimized Risk Indexing. *Applied Sciences*, 10(4), 4199. doi: 10.3390/app10124199.
- CONAPCI. (10 de septiembre 2020). Incendios urbanos en México [en línea]. <https://conapci.org/incendios-urbanos-en-mexico/>
- Di Pietro, R., Guarino, S., Verde, N. V., & Domingo-Ferrer, J. (2014). Security in wireless ad-hoc networks – A survey. *Computer Communications*, 51, 1-20. <https://doi.org/10.1016/j.comcom.2014.06.003>.
- Gamaleldin, A., Atef, A., Saker, H., & Shaheen, A. (2018). Fire Dataset. Outdoor-fire images and non-fire images for computer vision tasks. Kaggle. <https://www.kaggle.com/datasets/phylake1337/fire-dataset> [available].

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2013). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Gómez Mena, C. (27 de abril 2022). Quemaduras, tercer lugar como causa de muerte de niños en México [en línea]. <https://www.jornada.com.mx/notas/2022/04/27/sociedad/ocupan-las-quemaduras-el-tercer-lugar-de-muertes-en-infantes-en-el-pais/>.

Guo, K., Yang, P., Guo, D., & Liu, Y. (2019). Gas leakage monitoring with mobile wireless sensor networks. *Procedia Computer Science*, 154, 430-438. doi: 10.1016/j.procs.2019.06.061.

Iqbal, M., Setianingsih, C., & Irawan, B. (2020). Deep Learning Algorithm for Fire Detection. *10th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*, 237-242, doi: 10.1109/EECCIS49483.2020.9263456.

Kothandapani, V., Teja, V., Kumar, S., Divya, P., & D., B. (2021). Smart gas leak detection. *International Journal of Scientific Research in Science and Technology*, 8, 952-956.

Kou, L., Wang, X., Guo, X., Zhu, J., & Zhang, H. (2021). Deep learning based inverse model for building fire source location and intensity estimation. *Fire Safety Journal*, 121. <https://doi.org/10.1016/j.firesaf.2021.103310>.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., & Reed, S. (2015). SSD: Single shot multibox detector.

López-Molina, A., Vázquez-Román, R., & Díaz-Ovalle, C. (2012). Aprendizajes del Accidente de San Juan Ixhuatepec-México. *Información tecnológica*, 23(6), 121-128. <https://dx.doi.org/10.4067/S0718-07642012000600013>.

Lule, E., Mikeka, C., Ngenzi, A., & Mukanyiligira, D. (2020). Design of an IoT-based fuzzy approximation prediction model for early fire detection to aid public safety and control in the local urban markets. *Symmetry*, 12(9), 1391.

Meidelfi, D., Moodutor, H., Sukma, F., & Adnin, S. (2022). Android based spark and gas leak detection and monitoring. *Journal of Computer Networks, Architecture and High Performance Computing*, 4, 148-157. doi: 10.47709/cnahpc.v4i2.1489.

Mendoza, E., Fuentes, P., Benítez, I., Reina, D., & Núñez, J. (2020). Red de sensores inalámbricos multisalto para sistemas domóticos de bajo costo y área extendida. *Revista Iberoamericana de Automática e Informática industrial*, 17(4), 412-423.

Muhammad, K.; Ahmad, J.; Mehmood, I.; Rho, S.; and Baik, S. W. (2018). Convolutional Neural Networks Based Fire Detection in Surveillance Videos. *IEEE Access*, vol. 6, 18174-18183, doi: 10.1109/ACCESS.2018.2812835.

Mwedzi, N. A., Nwulu, N. I., and Gbadamosi, S. L. (2019). Machine Learning Applications for Fire Detection in a Residential Building. *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 1-4, doi: 10.1109/ICETAS48360.2019.9117318.

Nazir, A.; Mosleh, H.; Takruri, M.; Jallad, A.-H.; Alhebsi, H. (2022) Early Fire Detection: A New Indoor Laboratory Dataset and Data Distribution Analysis. *Fire*, 5(11). <https://doi.org/10.3390/fire5010011>.

Park, M., & Ko, B. C. (2020). Two-step real-time night-time fire detection in an urban environment using Static ELASTIC-YOLOv3 and Temporal Fire-Tube. *Sensors*, 20(8), 2202.

Pavez, F. A. (2021). Manual de autoprotección contra incendios para un emprendimiento unipersonal de producción artesanal(Bachelor's thesis).

Raj, S., Ankit, T., Rishabh, S., & Saurabh, S. (2021). Gas leak detection and smart alerting using IOT. *I-manager's Journal on Information Technology*, 10 (12). doi: 10.26634/jit.10.3.15052.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *IEEE 2016 conference on computer vision and pattern recognition (CVPR)*, p. 779-788. 10.1109/CVPR.2016.91.

Rostami, A.; Shah-Hosseini, R.; Asgari, S.; Zarei, A.; Aghdami-Nia, M.; Homayouni, S. (2022) Active Fire Detection from Landsat-8 Imagery Using Deep Multiple Kernel Learning. *Remote Sensors*, 14 (992). <https://doi.org/10.3390/rs14040992>.

Rzajj, W. & Al-Obaidi, B. (2022). Evaluation of a fire safety risk prediction model for an existing building. *Journal of the Mechanical Behavior of Materials*, 31(1), 64-70. <https://doi.org/10.1515/jmbm-2022-0007>.

Saponara, S., Elhanashi, A., & Gagliardi, A. (2021). Real-time video fire/smoke detection based on CNN in antifire surveillance systems. *J Real-Time Image Proc*, 18, 889–900. <https://doi.org/10.1007/s11554-020-01044-0>.

Secretaría de Salud (27 de abril 2021). Cuenta México con centro de alta especialidad para atender quemaduras graves [en línea]. <https://www.gob.mx/salud/prensa/170-cuenta-mexico-con-centro-de-alta-especialidad-para-atender-quemaduras-graves>.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv* 1409.1556.

Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, 61, 102332.

Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2020). Efficient processing of deep neural networks. *Synthesis Lectures on Computer Architecture*, 15(2), 1-341.

Thomazini, D., & de Albuquerque, P. U. B. (2020). *Sensores industriais: fundamentos e aplicações*. Saraiva Educação SA.

Wang, K.; Zhang, Y.; Wang, J.; Zhang, Q.; Chen, B. and Liu, D. (2018). Fire Detection in Infrared Video Surveillance Based on Convolutional Neural Network and SVM. 2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP), 162-167, doi: 10.1109/SIPROCESS.2018.8600510.

Wang, Z., Zhang, T., Wu, X., & Huang, X. (2022). Predicting transient building fire based on external smoke images and deep learning. *Journal of Building Engineering*, 47. <https://doi.org/10.1016/j.jobe.2021.103823>.

Yang, Z.; Bu, L. T., Wang, J., Ouyang, P., & Yuan, P. (2018). Fire Alarm for Video Surveillance Based on Convolutional Neural Network and SRU. 2018 5th International Conference on Information Science and Control Engineering (ICISCE), 232-236, doi: 10.1109/ICISCE.2018.00056.

Yang, Z.; Bu, L.; Wang, T.; Yuan, P.; Jineng, O. (2020). Indoor Video Flame Detection Based on Lightweight Convolutional Neural Network. *Pattern Recognition and Image Analysis*, 30, 551-564. doi: 10.1134/S1054661820030293.

Zhang, Y., Geng, P., Sivaparthipan, C. B., & Muthu, B. A. (2021). Big data and artificial intelligence based early risk warning system of fire hazard for smart cities. *Sustainable Energy Technologies and Assessments*, 45, 100986.



Esta obra está bajo una licencia de Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Explorando el Potencial de la WBB Modificada: Un Estudio Preliminar en la Evaluación del Equilibrio y la Prevención de Caídas en Entornos Clínicos

Exploring the Potential of Modified WBB: A Preliminary Study on Balance Assessment and Fall Prevention in Clinical Setting

Azucena Eunice Jiménez-Corona^a

azucena_jimenez@uaeh.edu.mx

Estefanía Olivares-Palacios^b

ol397073@uaeh.edu.mx

Aislinn Joally Ramírez-Méndez^b

ra397373@uaeh.edu.mx

Marisa Bautista-Lara^b

ba399821@uaeh.edu.mx

Domínguez Vega Zeus Tlaltecutili^c

ztdominguezv@uaemex.mx

Rigoberto Martínez-Méndez^c

rmartinezme@uaemex.mx

^a Autor de correspondencia: Universidad Autónoma del Estado de Hidalgo, Escuela Superior de Huejutla.

^b Universidad Autónoma del Estado de Hidalgo, Escuela Superior de Huejutla, Lic. Enfermería.

^c Universidad Autónoma del Estado de México, Facultad de ingeniería.

Resumen

La prevención de caídas, fuente principal de lesiones y mortalidad en adultos, se puede mejorar por medio de una adecuada evaluación del equilibrio. No obstante, el alto costo y complejidad de las herramientas actuales limitan su aplicación, particularmente en clínicas de atención primaria. Surge así la necesidad de metodologías más accesibles y económicas para optimizar el cuidado de salud en este segmento poblacional.

En el presente artículo se introduce un estudio que explora la aplicabilidad de una versión mejorada de la Wii Balance Board (WBBm). Este dispositivo, de bajo costo y portabilidad, ha sido respaldado por diversos investigadores en términos de validez. La muestra consistió en 95 adultos, con edades comprendidas entre 40 y 85 años, quienes fueron evaluados con la WBBm con el fin de correlacionar mediciones del Centro de Presión (CoP) con parámetros clínicos y antropométricos.

La investigación reveló que la WBBm proporciona señales efectivas para una evaluación cuantitativa del equilibrio y su uso es intuitivo y expedito. Sin embargo, se destaca que la interpretación del CoP está influenciada por múltiples variables clínicas y antropométricas. Esto sugiere la necesidad de cautela al generalizar la aplicabilidad de estas técnicas en una población diversa.

Para concluir, aunque queda camino por recorrer en esta área de investigación, la WBBm emerge como una herramienta con un alto potencial. Su adaptabilidad y practicidad pueden facilitar una valoración más eficiente del equilibrio en entornos clínicos de atención primaria.

Palabras clave: *equilibrio, adultos mayores, riesgo de caídas, enfermedades crónicas, centro de presión, plataforma biomecánica*

Abstract

Fall prevention, a major source of injury and mortality in adults, can be improved through proper balance assessment. However, the high cost and complexity of current tools limit their application, particularly in primary care clinics. There arises the need for more accessible and economical methodologies to optimize healthcare in this population segment.

In the present article, we introduce a study that explores the applicability of an enhanced version of the Wii Balance Board (WBBm). This low-cost and portable device has been supported by various researchers in terms of validity. The sample consisted of 95 adults, aged between 40 and 85, who were evaluated with the WBBm in order to correlate Center of Pressure (CoP) measurements with clinical and anthropometric parameters.

The research revealed that the WBBm provides effective signals for a quantitative assessment of balance and its use is intuitive and expedient. However, it is noted that the interpretation of the CoP is influenced by multiple clinical and anthropometric variables. This suggests the need for caution when generalizing the applicability of these techniques to a diverse population.

In conclusion, although there is still work to be done in this research area, the WBBm emerges as a tool with high potential. Its adaptability and practicality can facilitate a more efficient assessment of balance in primary care clinical settings.

Keywords: *Balance, Older Adults, Fall Risk, Chronic Diseases, Center of Pressure, Biomechanical Platform*

1 Introducción

La evaluación del equilibrio humano es un tema crítico en la salud y bienestar de individuos, especialmente en la población de adultos mayores. La pérdida de equilibrio y las caídas relacionadas son problemas comunes y graves en esta población, lo que lleva a una alta incidencia de lesiones y una disminución de la calidad de vida, (Center for Disease Control and Prevention, 2023; Gerardo et al., 2018; Rubenstein, 2006).

La obesidad y otras condiciones crónicas, así como el envejecimiento, también han sido relacionadas con un mayor riesgo de pérdida de equilibrio y caídas (Bruce et al., 2015; Fahimfar et al., 2022; Martínez, 2018). Estos factores pueden afectar negativamente la estabilidad postural y aumentar el riesgo de caídas, resultando en graves consecuencias para la salud como fracturas y mayor dependencia en las actividades diarias.

En este sentido, la medición precisa y efectiva del equilibrio se convierte en un aspecto esencial para la prevención, tratamiento, y rehabilitación en diversos campos médicos y biomecánicos (García Tirado, 2015; Peydro de Moya et al., 2005). A pesar de su importancia, los métodos convencionales para medir el equilibrio pueden ser costosos, invasivos, o inaccesibles en ciertos entornos. Las plataformas de fuerza tradicionales han sido ampliamente utilizadas, pero su alto costo limita su aplicación en entornos clínicos y de investigación más accesibles (Díaz & Mancilla, 2017).

Para profesionales de la salud, tales como médicos y enfermeras, un dispositivo con estas características reviste gran valor. Facilita una evaluación más ágil y precisa del riesgo de caídas, posibilitando intervenciones preventivas más acertadas. La monitorización en tiempo real del equilibrio y estabilidad puede ser una herramienta clave en rehabilitación y terapia personalizada, potenciando la eficacia del tratamiento y optimizando los recursos sanitarios. Hospitales en España, como el Hospital Nisa Sevilla Aljarafe, el Hospital Nisa Valencia al Mar y la Universidad Politécnica de Valencia, han reportado experiencias positivas en este ámbito. En particular, han implementado un programa para la Wii denominado eBaVIR, que combina la plataforma con ejercicio virtual para la rehabilitación de pacientes con daño cerebral adquirido. Esta solución ha demostrado ser una alternativa viable a la fisioterapia convencional, logrando resultados comparables (Gil-Gómez et al., 2011).

En este contexto, la Wii Balance Board (WBB) ha surgido como un instrumento prometedor para la evaluación del equilibrio. Pese a su diseño original orientado a videojuegos, diversos estudios han validado la WBB como un medio fiable y efectivo para medir el equilibrio en variadas poblaciones (Clark et al., 2010; Estevez Pedraza, 2017; Huurnink et al., 2013).

El propósito de este artículo es presentar una adaptación en hardware y firmware de la WBB, transformándola en un dispositivo portátil y asequible para evaluar el equilibrio. Esta adaptación se correlaciona con variables antropométricas en distintos contextos. La principal innovación de esta investigación radica en la reconfiguración de una tecnología ya existente hacia un instrumento eficiente y económico, con aplicaciones relevantes en áreas como la rehabilitación física, investigación biomecánica y entrenamiento deportivo. Esta adaptación ofrece una solución alternativa y efectiva frente a los métodos convencionales de medición del equilibrio.

En la sección de metodología se detallará el diseño del dispositivo, la elección de los participantes y los métodos de análisis aplicados, incluyendo referencias pertinentes en estas áreas

2 Metodología

2.1 Diseño del Estudio

Este estudio se llevó a cabo con un enfoque observacional, descriptivo y transversal, con el objetivo principal de caracterizar la población en términos de equilibrio y correlacionar estos datos con varios indicadores antropométricos y de salud.

La naturaleza observacional del estudio permitió una exploración detallada y no invasiva de las características de equilibrio en la muestra seleccionada. Al utilizar la Wii Balance Board modificada (WBBm), se buscó evaluar la funcionalidad de esta herramienta en un entorno real, analizando su aplicabilidad y eficacia en comparación con métodos convencionales de evaluación del equilibrio.

Para garantizar una evaluación integral, se implementaron herramientas de evaluación validadas, como el Test up and Go (TUG) y la prueba de Romberg, en conjunción el uso de la WBBm (Bruce et al., 2015). Estos métodos proporcionaron una base para analizar la capacidad del cuerpo para mantener el equilibrio y registrar el desplazamiento del centro de presión (CoP).

El diseño transversal permitió la recolección simultánea de información sobre la postura, el equilibrio, y las variables de salud y estilo de vida relacionadas en un momento específico. Esto facilitó una representación instantánea de la población, contribuyendo a una comprensión más profunda de la interrelación entre estas variables en un entorno clínico.

En resumen, el diseño de este estudio proporciona una estructura metódica y replicable, centrada en la exploración y caracterización de la población en términos de equilibrio y su relación con las variables antropométricas y de salud. La metodología empleada subraya la utilidad de las tecnologías accesibles y económicas en la evaluación del equilibrio, lo que puede tener aplicaciones significativas en campos como el diagnóstico clínico, el seguimiento, la rehabilitación física, la investigación biomecánica y el entrenamiento deportivo.

2.2 Participantes del estudio

En el estudio participaron adultos de entre 40 y 85 años, derechohabientes de la Clínica Hospital ISSSTE Huejutla, con antecedentes de caídas o accidentes relacionados con el equilibrio y con normopeso, sobrepeso u obesidad. Los criterios de exclusión fueron específicos y se detallan en la Tabla 1.

Tabla 1 Criterios de inclusión y exclusión

Criterios	Descripción
Inclusión	Adultos (hombres o mujeres) mayores de 40 a 85 años
	Ser derechohabiente de la Clínica Hospital ISSSTE Huejutla
	Antecedentes de caídas o accidentes relacionados con el equilibrio
	Normopeso, sobrepeso u obesidad
	Condiciones de vida no idóneas
Exclusión	No tener la edad de 40 a 85 años
	No ser derechohabiente de la Clínica Hospital ISSSTE Huejutla
	Enfermedades ortopédicas o neurológicas
	Consumo de alcohol

Reclutamiento y Consentimiento

Se sensibilizó a la población de derechohabientes de la Clínica Hospital ISSSTE-Huejutla a través de pláticas, folletos y trípticos, explicando el objetivo del estudio (ver Fig. 1). Todos los participantes proporcionaron su consentimiento informado antes de la participación en el estudio, en cumplimiento con la confidencialidad de acuerdo con el artículo 6 inciso A y 16°, y se les entregó una copia de este. El protocolo de la investigación fue revisado y aprobado por el Comité de Ética e Investigación del Instituto de Ciencias de la salud (Oficio Comiteei.icsa 023/2022) y de la Clínica-Hospital del ISSSTE de Huejutla, Hidalgo.



Fig. 1 Invitación a participar en el proyecto, explicándoles el objetivo del estudio.

La población evaluada consistió en 95 participantes, incluyendo 32 hombres y 63 mujeres, con edades, pesos y tallas que abarcaron un amplio rango. (Véase la Tabla 2 para una descripción detallada de las características antropométricas de la muestra).

Tabla 2 Características antropométricas de la muestra poblacional participante.

<i>Característica</i>	<i>Hombres rango, (promedio $\pm$ SD)</i>	<i>Mujeres rango, (promedio $\pm$ SD)</i>	<i>Total rango, (promedio $\pm$ SD)</i>
<i>Número</i>	32	63	95
<i>Edad (años)</i>	41-83 (61.12 $\pm$ 8.32)	48-75 (63.81 $\pm$ 7.26)	41-83 (59.75 $\pm$ 8.55)
<i>Peso (kg)</i>	35.40-112.50 (66.67 $\pm$ 12.97)	49.00-92.50 (72.37 $\pm$ 9.88)	35.40-112.50 (63.78 $\pm$ 13.46)
<i>Talla (m)</i>	1.30-1.83 (1.56 $\pm$ 0.08)	1.50-1.83 (1.63 $\pm$ 0.06)	1.30-1.69 (1.52 $\pm$ 0.07)

La mayoría de los participantes son de la comunidad de Huejutla y sus alrededores, incluyendo personas dedicadas al campo, amas de casa, docentes, secretarías, pensionados y jubilados. Esta selección proporciona una muestra representativa de la población adulta de la región, enfocada en individuos con antecedentes de caídas o problemas de equilibrio, en un rango de edad de 40 a 85 años.

2.3 Instrumentación

Para utilizar la estructura de la WBB (Wii Balance Board), se llevó a cabo una modificación en su circuitería, con el objetivo de asegurar una frecuencia de muestreo estable a 50 Hz sin pérdida de datos. Esta modificación implicó una mejora en la tarjeta principal de la tarjeta, permitiendo controlar los convertidores analógico-digitales (ADC) mediante el ESP32 para leer los datos de los sensores de la plataforma (Joan, 2017/2017). Los detalles de estas modificaciones se pueden apreciar en la Fig. 2.

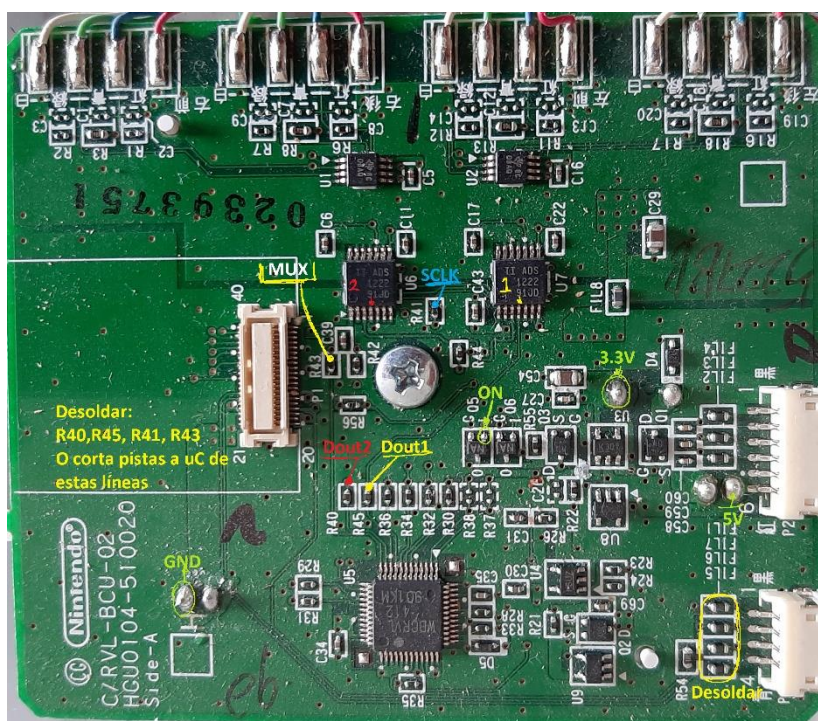


Fig. 2 Modificaciones a la tarjeta original de la WBB, basadas en (Joan, 2017).

Tras realizar las modificaciones necesarias, se conectaron los componentes a la tarjeta ESP32. Simultáneamente, se diseñó un circuito electrónico para controlar una pantalla OLED, botones y un buzzer, creando así una interfaz de usuario. Se incorporó un módulo RTC para registrar la fecha y hora de cada prueba, facilitando la identificación de los pacientes. Cada archivo de datos se almacena con la fecha y hora en una tarjeta microSD. La Fig. 3 ilustra los principales componentes electrónicos y las modificaciones hechas a la WBB.

La estructura y los sensores no sufrieron cambios, manteniendo así la capacidad de la WBB para adquirir señales del Centro de Presión (CoP) de manera similar a las plataformas comerciales, tal como se validó anteriormente (Clark et al., 2010). Además, se mejoró su desempeño al garantizar una frecuencia de muestreo estable.

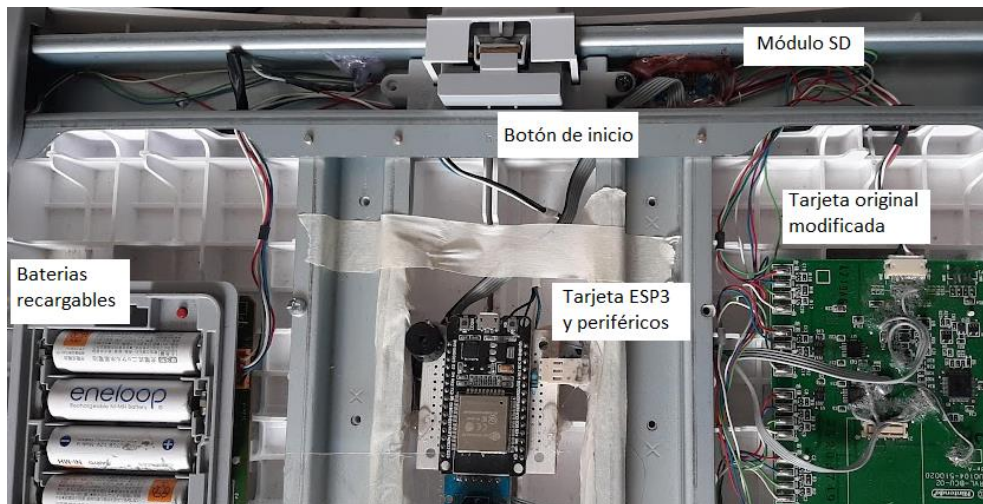


Fig. 3 Modificaciones y circuitos de control de la WBBm.

En la Fig. 4 se muestra la plataforma en su presentación final. Se puede observar que el tamaño se mantiene sin cambios, así como la estructura, por lo que las pruebas realizadas por (Clark et al., 2010), donde asegura la validez y confiabilidad de la plataforma, se pueden considerar válidas también para esta versión. Una versión similar, pero empleando un tipo diferente de pantalla, diferente microcontrolador y enfocada en la evaluación de riesgo de caída fue publicada en (Estevez Pedraza, 2017).



Fig. 4 Presentación final de la WBBm con sus partes principales.

La Fig. 5 presenta un diagrama de flujo simplificado del firmware programado en la tarjeta ESP32 para el control de la WBB. El procedimiento para el paciente es similar al uso de una báscula de baño. El paciente debe pararse sobre la plataforma y seguir las indicaciones visuales y auditivas para completar una prueba de 80 segundos. La prueba puede realizarse rápidamente, sin tomar más de 2 minutos del tiempo de una consulta normal en cualquier centro de atención de salud primario.

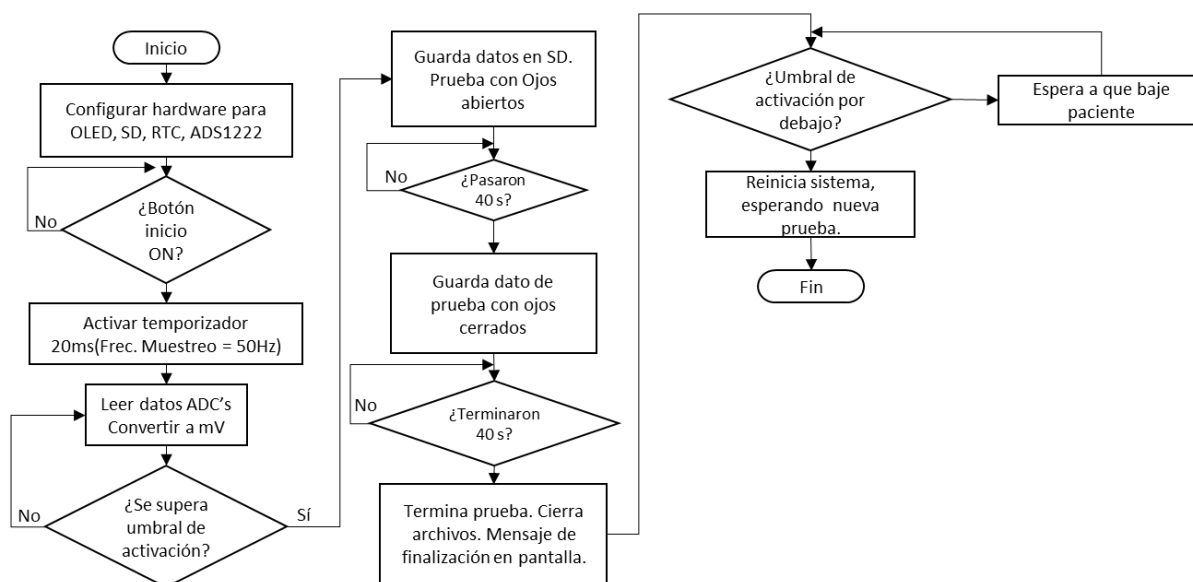


Fig. 5 Diagrama de flujo del firmware programado en la tarjeta ESP32 para el control de la WBBm.

En conclusión, esta plataforma de bajo costo podría simplificar la recolección de datos cuantitativos relacionados con el equilibrio de los pacientes. La información recopilada podría ser valiosa tanto para el personal médico como para los investigadores interesados en desarrollar modelos de predicción para caídas o equilibrio, y su relación con parámetros físicos o neurológicos individuales

2.3 Protocolo de recopilación de datos clínicos y de equilibrio

Como ya se mencionó en la subsección Participantes del estudio, se invitó libremente a los pacientes de la Clínica Hospital ISSSTE-Huejutla para participar en el estudio. A todos los 95 participantes incluidos en el estudio se les aplicó el siguiente protocolo:

2. 3.1 Medición de datos clínicos

1. Signos Vitales: Medición de signos vitales, incluyendo la presión arterial utilizando el Kit Simple Baumanómetro Anaeroide Eco Negro CheckAteck.
2. Medición Anatómica: Medición de circunferencia de abdomen, brazo y muñeca mediante el uso de una cinta métrica Lufkin (Executive Thinline).
3. Composición Corporal y Peso: Utilizando el sistema comercial TANITA FitScan (Modelo BC-730F), se obtuvieron parámetros como índice de grasa corporal, índice de masa ósea, porcentaje de agua, etc.
4. Estado Emocional: Evaluación utilizando la escala CES-D7 (Center for Epidemiologic Studies Depression Scale-7) (Okun et al., 1996).
5. Índice de Comorbilidad de Charlson: Puntuación basada en patologías diagnosticadas (Charlson et al., 2022).
6. Cuestionario de Caídas: Enfocado en caídas experimentadas en el último año (Tinetti et al., 1986).
7. Índice de Barthel: Utilizado para evaluar la capacidad de un individuo para realizar actividades básicas de la vida diaria (AVD) (Bernabeu-Wittel et al., 2019).

La Fig. 6 muestra a una participante del estudio y al personal de salud mientras realizan la toma de datos clínicos.



Fig. 6 Recopilando los datos clínicos.

2. 3. 2 Evaluación de Equilibrio

1. Ejercicios Iniciales: Incluyendo levantamiento y sentado de una silla, y ejercicios de equilibrio en diversas posiciones.
2. Evaluación de la Marcha: Evaluación del desempeño físico en una distancia de 4.6 metros.
3. Test de Estación Unipodal: Prueba específica para adultos mayores para prevenir el riesgo de caídas.
4. Prueba de Romberg: Utilizando una plataforma biomecánica para medir el Centro de Presión (CoP) (Gallamini et al., 2021).

La Fig. 7 muestra una de las participantes realizando la prueba de Romberg.



Fig. 7 Paciente realizando la prueba de Romberg utilizando la plataforma WBBm.

2. 3. 3 Evaluación de Bioimpedancia

El monitor de composición corporal TANITA FitScan Modelo BC-730F se utilizó para medir diversas variables corporales utilizando el método de bioimpedancia. Es importante mencionar que las personas con marcapasos no deben hacer uso de este dispositivo.

Además de las pruebas de equilibrio, se recopilaban datos sobre diversas variables, incluyendo edad, sexo, presión arterial en dos posiciones, frecuencia cardíaca, frecuencia respiratoria, SPO₂, IMC y medidas de bioimpedancia.

Este enfoque integral permite una visión completa de la salud y el estado nutricional de los individuos, y las evaluaciones realizadas pueden ser útiles para monitorear los cambios en la capacidad funcional a lo largo del tiempo y para determinar la necesidad de asistencia o intervenciones terapéuticas.

2.4 Análisis de datos

En el marco de esta investigación, la prioridad se estableció en el desarrollo de la plataforma WBBm y la implementación del protocolo de recolección de datos, dejando un análisis minucioso de la información para posteriores etapas. Los datos obtenidos se estructuraron y almacenaron en una base específica y se procesaron mediante el software Matlab 2023a para un análisis exploratorio inicial.

Se empleó Matlab 2023a, un entorno de programación de alto nivel, para el procesamiento y visualización de datos. Se generaron gráficos que proporcionan una mirada inicial a la información recolectada, útiles para discernir tendencias, patrones y posibles correlaciones entre variables.

Además, se estimaron métricas de evaluación del equilibrio previamente descritas por (Quijoux et al., 2021). Dichas métricas son ampliamente reconocidas en el análisis y evaluación del equilibrio, siendo esenciales para el diagnóstico y evaluación de pacientes con problemas de balance. También se efectuó una comparación entre los resultados obtenidos en este estudio y los reportados por (Santos & Duarte, 2016). Lo anterior con el fin de comparar si los resultados obtenidos en las mediciones realizadas en este estudio son comparables con otros métodos convencionales, como las plataformas de fuerza.

La velocidad del CoP es una de las métricas derivadas más empleadas para evaluar el equilibrio (Quijoux et al., 2021). Por ello, se optó por esta métrica para un análisis estadístico adicional. Con el objetivo de explorar las diferencias entre quienes han sufrido caídas y quienes no en el conjunto de datos del estudio, se realizó una prueba t de dos muestras. Para investigar la influencia de la edad en el análisis, se aplicó un ANCOVA para detectar diferencias en los dos grupos, considerando la edad como covariable.

Es pertinente mencionar que el análisis exhaustivo de la base de datos y la potencial creación de modelos predictivos entre las variables exceden el propósito del presente artículo. Se prevé un análisis más detenido en futuras publicaciones, donde se ahondará en la interpretación y modelado de la información.

Para concluir, aunque el análisis presentado en este artículo es preliminar, la estructuración y recolección de datos sientan las bases para futuras investigaciones y análisis de mayor profundidad."

3 Resultados

Para validar la precisión del dispositivo desarrollado en este trabajo, se llevó a cabo una comparación de los promedios obtenidos de diversas variables del CoP con resultados de dos estudios anteriores. Quijoux et al., en 2021, condujeron una revisión de las métricas derivadas del CoP más frecuentemente empleadas en la evaluación del equilibrio en adultos mayores (133 participantes mayores de 60 años). Este estudio reportó valores promedio usando una WBB (Quijoux et al., 2021). Paralelamente, Santos et al., en 2016, efectuaron mediciones empleando una plataforma de fuerza (OPT400600-1000, AMTI) en un conjunto de 163 sujetos con edades comprendidas entre los 18 y 85 años, y posteriormente publicaron su base de datos (Santos & Duarte, 2016).

Para determinar la congruencia entre los datos obtenidos en este trabajo y los mencionados anteriormente, se calcularon idénticas métricas (41 en total, tal como se reportan en (Quijoux et al., 2021) y se representaron gráficamente. La Figura 8 ilustra que, si bien existen diferencias en las amplitudes, todas las variables siguen tendencias similares.

Un análisis de correlación reveló un coeficiente superior a $r=0.97$ al comparar las variables de este estudio con las de Santos et al. y Quijoux et al. Esto sugiere que, a excepción de las amplitudes, los valores son consistentes entre los estudios. Las discrepancias en las amplitudes podrían atribuirse a la diversidad de la población examinada en este estudio o a la sensibilidad del instrumento utilizado. Esta particularidad amerita un análisis más profundo en investigaciones subsecuentes.

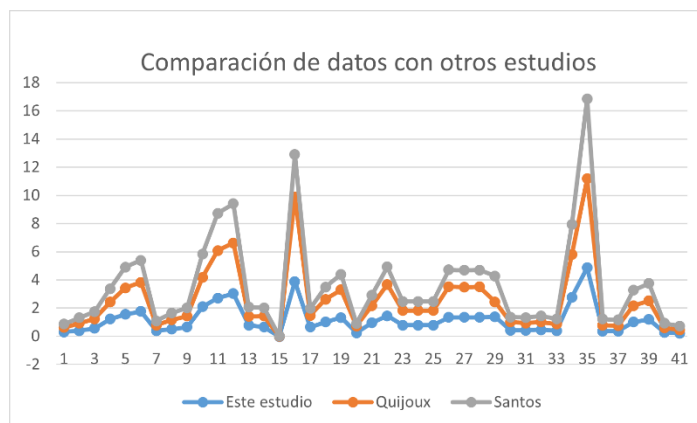


Fig. 8 Comparación de los valores obtenidos con la WBBm y los datos obtenidos por otros investigadores. El eje vertical está en unidades en cm. En el eje horizontal, cada numeral indica una de las variables analizadas en esos estudios, todas derivadas del CoP.

A continuación, se presentan los resultados preliminares del análisis de datos obtenidos mediante la plataforma WBBm, centrados en el estudio de las señales de desplazamiento medio-lateral (ML) y antero-posterior (AP). Las figuras a continuación representan las observaciones de uno de los participantes.

En la Fig. 9, se muestran las señales de desplazamiento del Centro de Presión (CoP) comparando las condiciones con ojos abiertos (OA) y con ojos cerrados (OC) durante la prueba de Romberg.

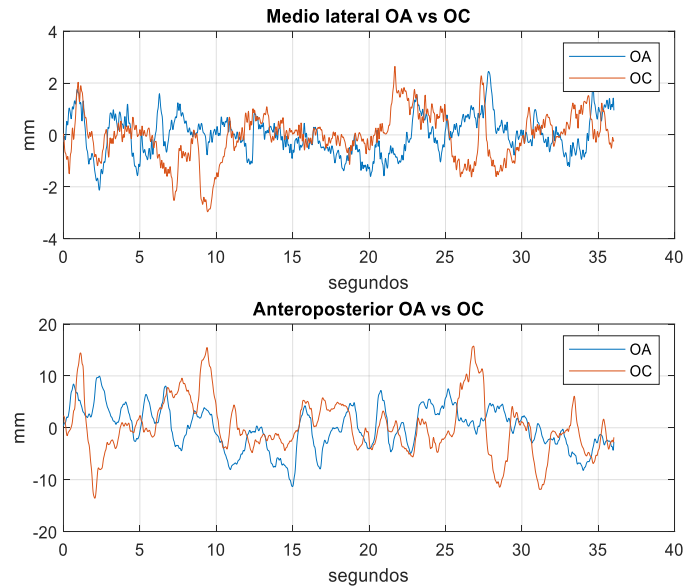


Fig. 9 Señales de desplazamiento del CoP obtenidas de un participante.

Para este participante, la frecuencia de la señal aumenta notablemente con los ojos cerrados, lo cual es un indicativo claro del efecto en el equilibrio. Así mismo, se pueden ver amplitudes de señal más grandes con los ojos cerrados. Esta tendencia se suele representar mejor mediante estabilogramas. En la Fig. 10, se observa el estabilograma para las condiciones de ojos abiertos, con el de ojos cerrados sobrepuesto, revelando que el área de desplazamiento se incrementa cuando el participante cerró los ojos.

El análisis de estas señales, junto con su posterior análisis en Matlab, permitirá extraer valores característicos y cuantificables de cada participante. La correlación de estos datos con variables clínicas contribuirá al desarrollo de modelos predictivos basados en las mediciones del CoP. Estos modelos pueden ofrecer herramientas de apoyo valiosas en el diagnóstico y seguimiento de los pacientes.

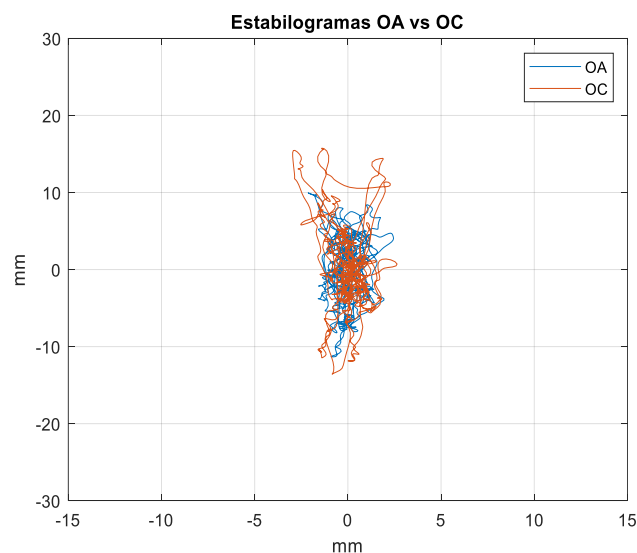


Fig. 10 Estabilograma comparativo entre ojos cerrados y ojos abiertos.

Dentro de la medición de datos clínicos, se les preguntó sobre el número de caídas que habían tenido en el último año. La literatura menciona que una persona con más caídas suele presentar un decrecimiento en su equilibrio y eso debería poderse evaluar utilizando las mediciones derivadas del CoP. Un meta análisis sobre este tema sugiere el cálculo del área de balanceo con los ojos cerrados (AREACCOC) como uno de los parámetros más relacionados con esto y que podría discriminar entre los participantes que se han caído una o más veces (Kozinc et al., 2020) .

Con el fin de probar esto se separaron los participantes en tres grupos (0 caídas; 1 caída; 2 o más caídas. Y se calculó el promedio de valor del área de su balanceo con ojos cerrados. El resultado se puede observar en la Fig. 11. No se ve una relación como la esperada (mayor valor de AREACCOC a mayor número de caídas), además, la variabilidad de los datos es muy grande. Por lo que estadísticamente no hay diferencia entre los grupos.

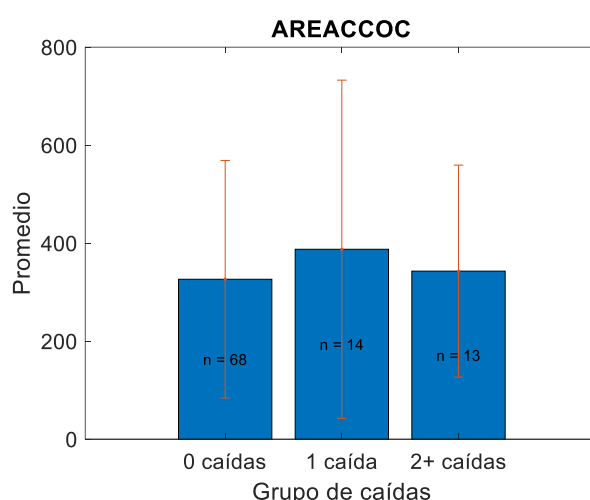


Fig. 11 Valores promedio y SD del valor del área de balanceo (mm²), separados en grupos por número de caídas en el último año.

Una gráfica de dispersión de datos del valor del AREACCOC con respecto a la edad, muestra con más detalle la gran dispersión de datos, ver Fig. 12.

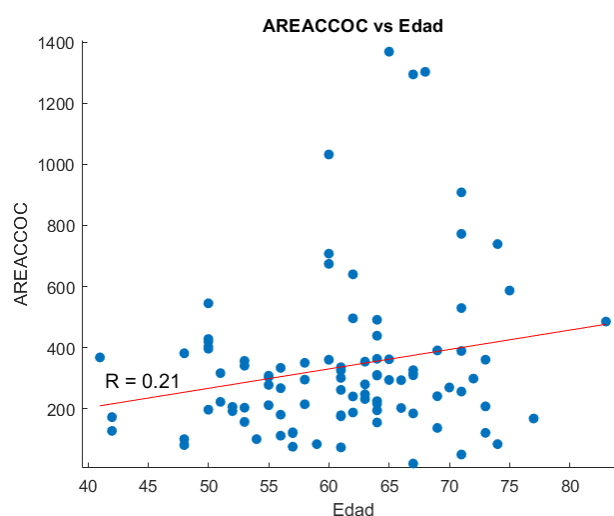


Fig. 12 Gráfica de dispersión de datos de AREACCOC vs Edad.

Hay dos factores que podrían explicar esta discrepancia con el estado del arte.

1. El rango de edad empleado en este estudio es de 41 a 85 años. En el estado del arte se mencionan grupos 65 años hacia arriba.

2. El grupo en este estudio son personas con diversas enfermedades, ya sean crónicas o temporales. En la Fig. 12 se observa la distribución de los participantes por número de enfermedades crónicas que padecen y divididos por sexo. Mientras que, en el estado del arte, generalmente son grupos homogéneos, relativamente sanos o no mencionan esas características.

Fig. 13 Cantidad de participantes con enfermedades crónicas, por sexo.

En el marco de la presente investigación, se trabajó con una muestra poblacional caracterizada por su notable heterogeneidad en aspectos antropométricos, clínicos y etarios. Detalles específicos de las características de esta muestra se detallan en la Tabla 3. Esta diversidad de sujetos introduce una complejidad adicional al estudio. A menudo, la literatura existente ha validado la posibilidad de determinar el riesgo de caídas o perturbaciones del equilibrio en cohortes más homogéneas, típicamente enfocándose en poblaciones geriátricas.

Tabla 3 Algunas características de la muestra participante en este estudio.

VARIABLE	N	%
Edad	95	100
Sexo		
Mujeres	62	65.26
Hombres	33	34.73
IMC (índice de masa corporal)		
Desnutrición	1	1.05
Normopeso	33	34.73
Sobrepeso	34	35.78
Obesidad grado I	19	20
Obesidad grado II	7	7.36
Obesidad grado III	1	1.05
Frecuencia cardiaca		
Bradicardia	6	6.31
Normal	89	93.68
Taquicardia	0	0
Frecuencia respiratoria		
Bradipnea	1	1.05
Eupnea	59	62.10
Taquipnea	35	36.84
Tensión arterial		
Presión arterial óptima	29	30.5
Presión arterial subóptima	53	55.8
Presión arterial límite	9	9.5
Hipertensión grado I	2	2.1
Hipertensión grado II	2	2.1
Hipertensión grado III	0	0
Bioimpedancia Tanita	91	95.78
Circunferencias corporales	23	24.21

No obstante, la variedad inherente a la muestra seleccionada para esta investigación brinda una oportunidad para construir una base de datos más amplia y diversa. Esta base será importante para evaluar, con mayor precisión, las capacidades predictivas de algoritmos y modelos propuestos en el estado del arte actual. Estos análisis e investigaciones serán objeto de futuras publicaciones.

Análisis ANCOVA

Dado que a través de un análisis estadístico descriptivo no se pudo observar diferencias significativas entre los grupos, se recurrió de manera preliminar al análisis ANCOVA (Análisis de Covarianza). El ANCOVA es una técnica estadística que combina los principios del ANOVA y la regresión. Es especialmente útil para ajustar o controlar la variabilidad de una o más variables covariables continuas que pueden estar influenciando la variable dependiente. De esta forma, ANCOVA permite evaluar las diferencias entre las medias de grupos ajustando simultáneamente los efectos de una o más covariables, lo cual puede mejorar la precisión y validez de las pruebas.

Sin embargo, antes de aplicar este tipo de análisis, se deben cumplir algunas consideraciones, entre ellas la normalidad de datos. De manera preliminar, ya que en futuros trabajos se realizará un análisis más exhaustivo, se realizó el análisis ANCOVA en la variable derivada del CoP llamada Velocidad Media en direcciones Medio-lateral (ML) y Anteroposterior (AP) (Quijoux et al., 2021). La distribución de datos no es normal, así que se realizó una transformación logarítmica para cumplir con este criterio. La transformación logarítmica fue seleccionada debido a su capacidad para estabilizar la varianza y reducir el sesgo en los datos con distribuciones asimétricas o con colas pesadas. Esta transformación es particularmente eficaz para datos que crecen exponencialmente o que presentan heterocedasticidad. En nuestro conjunto de datos, la transformación logarítmica demostró ser la más adecuada para satisfacer los supuestos de normalidad requeridos para el análisis ANCOVA.

En la Fig. 14 se observa como la transformación logarítmica aplicada consigue normalizar los datos. Una prueba Shapiro-Wilk corroboró la normalidad de los datos ($W=0.98$, $p=0.15$)

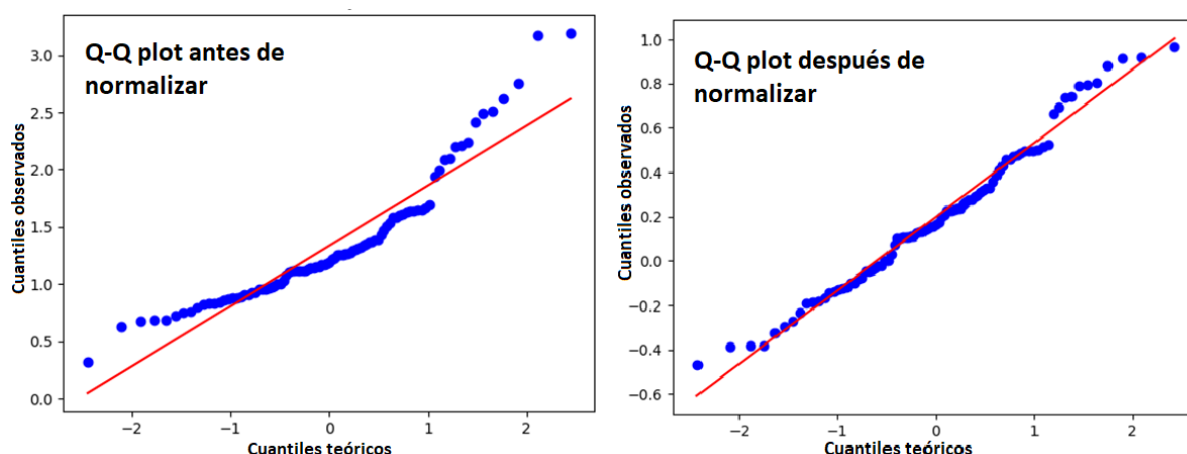


Fig. 14 Q-Q plot que demuestra la normalidad antes y después de la normalización.

A los datos ahora normalizados, se les aplicó una prueba t-student para comparar entre grupos de "fallers" (personas que se han caído al menos una vez en el último año) y "non fallers" (personas que no se han caído en el último año). Esta prueba arrojó ($F = 0.67$, $p = 0.41$) con lo que se puede inferir que no hay diferencias significativas entre los grupos. Sin embargo, los resultados de la prueba ANCOVA, usando la edad como covariable ($F = 3.74$, $p = 0.02$) muestran diferencias significativas entre las muestras. Este resultado está en línea con lo descrito en la literatura, donde se menciona que la edad es un factor determinante para la distinción de pacientes con y sin caídas. Para ambas pruebas se consideró un nivel de significancia $p > 0.05$.

4 Discusión

La correcta evaluación del estado de equilibrio es crucial para la prevención de caídas, reduciendo así significativamente los costos de atención médica. Entre 2013 y 2018 en México, se registraron 16,829 pacientes con fractura de cadera, generando elevados costos en atención médica (Pech-Ciau et al., 2021). Los gastos promedio en atención médica por caídas en hospitales para personas mayores de 65 años son considerables, alcanzando cifras de \$ 470,677,181.01 USD tan solo en Brasil, entre 2020 y 2022 (Lima et al., 2022). Este gasto representa una carga económica importante para los sistemas de salud pública, reforzando la necesidad de implementar medidas preventivas efectivas.

En este contexto, el estudio dirigió su atención en el desarrollo y la implementación de la plataforma WBBm para la recopilación de datos sobre el equilibrio humano. A pesar de no haber identificado correlaciones significativas en esta investigación preliminar, los resultados no refutan la posibilidad de su existencia; por el contrario, sugieren la necesidad de estudios más detallados y con muestras más amplias. El análisis mediante ANCOVA, al considerar otras covariables, soporta esta idea. A pesar de las discrepancias observadas respecto a la literatura, el estudio subraya la utilidad de la WBB modificada como una herramienta de evaluación del equilibrio accesible y económica. Tal como (Park & Lee, 2014) resaltaron, la WBB, al compararse con una Plataforma de fuerza de grado de laboratorio, mostró viabilidad a pesar de las diferencias en precisión y sensibilidad.

La investigación también reveló que, en el grupo de participantes, existe una notable prevalencia de enfermedades metabólicas, en particular Diabetes Mellitus e Hipertensión arterial. Estas afecciones tienen una implicación directa en el equilibrio, afectando la masa ósea, muscular y adiposa, así como los sistemas vestibulares y propioceptivos. (Corcuera-Ciudad et al., 2019) mencionó la asociación entre diabetes mellitus 2 y trastornos de marcha y equilibrio. La diabetes mellitus 2, debido a su impacto en el metabolismo muscular y la neuropatía periférica, incrementa considerablemente el riesgo de caídas en adultos mayores.

Incorporando hallazgos de estudios previos, se ha observado que la diferencia entre ambos pies en términos de estabilidad es mayor en posiciones con ojos abiertos que con ojos cerrados (Inglés et al., 2017). A la luz de estos resultados, es evidente la importancia del estabilograma para el análisis del equilibrio en personas mayores propensas a caídas. Además, debido a los cambios naturales en el sistema nervioso y el aparato locomotor durante el envejecimiento, es esencial comprender cómo afectan el equilibrio y la marcha (Lorena Cerda, 2014). La incorporación de técnicas de análisis más avanzadas, como el análisis espectral del estabilograma, podría proporcionar pistas adicionales sobre estos cambios (Smith, 2015). Se sugiere, en investigaciones futuras, explorar la relación entre medidas antropométricas y la relación del aumento de peso con el deterioro del equilibrio, ya que el incremento ponderal puede influir negativamente en la sensibilidad de los mecanorreceptores y en el ajuste del centro de presión, incrementando el riesgo de caídas (Hue et al., 2007).

Finalmente, las limitaciones del estudio, como el tamaño de la muestra y la falta de grupos control, junto con la diversidad de la muestra, evidencian la necesidad de una investigación más profunda y detallada. Sin embargo, los datos acumulados y organizados durante la investigación proporcionan un sólido punto de partida para futuras indagaciones y eventualmente, la implementación de programas de intervención basados en la WBB.

5 Conclusiones

El presente estudio logró adaptar con éxito la Wii Balance Board (WBB) modificada para la evaluación del equilibrio en contextos clínicos. Si bien no se detectaron correlaciones significativas entre el Centro de Presión (CoP) y ciertas variables relacionadas con la salud y el estilo de vida, los datos recopilados subrayan la intrincada naturaleza de dichas relaciones. Además, mediante la comparación con estudios anteriores, se ha confirmado que las medidas obtenidas con la WBB modificada son comparables a resultados de plataformas convencionales. Es esencial destacar la funcionalidad de la WBB modificada como herramienta de bajo costo en la evaluación del equilibrio.

Las caídas son un problema emergente en salud pública, impactando en mayor medida a individuos de la tercera edad. Ante esta situación, se vuelve imperativo sensibilizar a la sociedad sobre la relevancia de evaluar y fortalecer el equilibrio como medida preventiva. Esta investigación se posiciona como un esfuerzo para proporcionar una herramienta accesible que, a la vez, podría contribuir a una disminución de caídas, mejorando la calidad de vida y reduciendo gastos en atención médica.

La optimización de la WBB ha mostrado avances notables, principalmente en la estabilidad del muestreo y la mitigación de pérdida de datos, proponiendo una alternativa rentable frente a equipos de evaluación de equilibrio de mayor costo. La posibilidad de replicar tendencias observadas en otros estudios confirma el potencial de esta herramienta, lo cual podría democratizar el acceso a herramientas de evaluación, permitiendo su uso en diversos entornos y poblaciones.

Para futuras líneas de investigación, se sugiere expandir el alcance de la WBB hacia la rehabilitación física y en la valoración del equilibrio en pacientes con patologías específicas. También es crucial investigar las posibles razones para las diferencias de amplitud observadas en comparación con estudios previos. Asimismo, es esencial profundizar en el entendimiento de las relaciones entre equilibrio y diversas variables de salud y estilo de vida.

En suma, aunque este estudio no estableció correlaciones significativas, brinda un aporte valioso a la literatura especializada. Resalta el potencial de la WBB modificada en la evaluación del equilibrio y sienta las bases para investigaciones subsecuentes en este dominio.

6 Referencias

- Bernabeu-Wittel, M., Díez-Manglano, J., Nieto-Martín, D., Ramírez-Duque, N., Ollero-Baturone, M., Abella-Vázquez, L., Aguayo-Canela, M., Albert-Coll, M., Alemán, A., Aligué, J., Barbé Gil-Ortega, J., Barón-Franco, B., Bernabeu-Wittel, M., de la Calle-Cabrera, C., Cassani-Garza, M., Castillo-Blasco, M., Díez, F., Díez-Manglano, J., Escalera-Zalvide, A., ... Villegas-Bruguera, E. (2019). Simplification of the Barthel scale for screening for frailty and severe dependency in polypathological patients. *Revista Clínica Española (English Edition)*, 219(8), 433–439. <https://doi.org/10.1016/j.rceng.2019.04.011>
- Bruce, D., Hunter, M., Peters, K., Davis, T., & Davis, W. (2015). Fear of falling is common in patients with type 2 diabetes and is associated with increased risk of falls. *Age and Ageing*, 44(4), 687–690. <https://doi.org/10.1093/ageing/afv024>
- Center for Disease Control and Prevention. (2023, julio 19). *Facts About Falls*. <https://www.cdc.gov/falls/facts.html>
- Charlson, M. E., Carrozzino, D., Guidi, J., & Patierno, C. (2022). Charlson Comorbidity Index: A Critical Review of Clinimetric Properties. *Psychotherapy and Psychosomatics*, 91(1), 8–35. <https://doi.org/10.1159/000521288>

- Clark, R. A., Bryant, A. L., Pua, Y., McCrory, P., Bennell, K., & Hunt, M. (2010). Validity and reliability of the Nintendo Wii Balance Board for assessment of standing balance. *Gait & Posture*, 31(3), 307–310. <https://doi.org/10.1016/j.gaitpost.2009.11.012>
- Corcuera-Ciudad, R., Patiño-Villena, A. F., Paima-Olivari, R., Chamberg-Michilot, D., Parodi, J. F., Runzer-Colmenares, F. M., Corcuera-Ciudad, R., Patiño-Villena, A. F., Paima-Olivari, R., Chamberg-Michilot, D., Parodi, J. F., & Runzer-Colmenares, F. M. (2019). Trastornos de la marcha y el equilibrio en adultos mayores y su asociación con diabetes mellitus tipo 2. *Medicina interna de México*, 35(5), 676–684. <https://doi.org/10.24245/mim.v35i5.2554>
- Díaz, G., & Mancilla, S. (2017). Prototipo De Plataforma De Estabilometría Para El Análisis Del Equilibrio Corporal. *Memorias del Congreso Nacional de Ingeniería Biomédica*, 4(1), Article 1. <http://memoriascnib.mx/index.php/memorias/article/view/287>
- Estevez Pedraza, A. G. (2017). *Diseño y construcción de un dispositivo portátil para medición del centro de presión del cuerpo humano* [Maestría, Universidad Autónoma del Estado de México]. <http://ri.uaemex.mx/handle/20.500.11799/67806>
- Fahimfar, N., Yousefi, S., Noorali, S., Gharibzadeh, S., Sanjari, M., Khalagi, K., Mehri, A., Shafiee, G., Heshmat, R., Nabipour, I., Amini, A., Darabi, A., Heidari, G., Larijani, B., & Ostovar, A. (2022). The association of cardio-metabolic risk factors and history of falling in men with osteosarcopenia: A cross-sectional analysis of Bushehr Elderly Health (BEH) program. *BMC Geriatrics*, 22(1), 46. <https://doi.org/10.1186/s12877-021-02657-1>
- Gallamini, M., Piastra, G., Lucarini, S., Porzio, D., Ronchi, M., Pirino, A., Scoppa, F., Masiero, S., & Tognolo, L. (2021). Revisiting the Instrumented Romberg Test: Can Today's Technology Offer a Risk-of-Fall Screening Device for Senior Citizens? An Experience-Based Approach. *Life*, 11(2), Article 2. <https://doi.org/10.3390/life11020161>
- García Tirado, J. J. (2015). Análisis de los procesos de integración neurosensorial implicados en el control de la estabilidad postural en pacientes con disfunción cráneo-mandibular [Ph.D. Thesis, Universitat Internacional de Catalunya]. En *TDX (Tesis Doctorals en Xarxa)*. <https://www.tdx.cat/handle/10803/334979>
- Gerardo, G., Alemán, S., Manuel, V., Rodríguez, V., de Lourdes, M., Valdez, R., Selene, B., & Muñoz, Z. (2018). Factores asociados con caídas en el adulto mayor. *Paraninfo digital*, 28(XII), 1–3. <http://www.index-f.com/para/n28/pdf/e025.pdf>
- Gil-Gómez, J.-A., Lloréns, R., Alcañiz, M., & Colomer, C. (2011). Effectiveness of a Wii balance board-based system (eBaViR) for balance rehabilitation: A pilot randomized clinical trial in patients with acquired brain injury. *Journal of Neuroengineering and Rehabilitation*, 8, 30. <https://doi.org/10.1186/1743-0003-8-30>
- Hue, O., Simoneau, M., Marcotte, J., Berrigan, F., Doré, J., Marceau, P., Marceau, S., Tremblay, A., & Teasdale, N. (2007). Body weight is a strong predictor of postural stability. *Gait & Posture*, 26(1), 32–38. <https://doi.org/10.1016/j.gaitpost.2006.07.005>
- Huurnink, A., Fransz, D. P., Kingma, I., & van Dieën, J. H. (2013). Comparison of a laboratory grade force platform with a Nintendo Wii Balance Board on measurement of postural control in single-leg stance balance tasks. *Journal of Biomechanics*, 46(7), 1392–1395. <https://doi.org/10.1016/j.jbiomech.2013.02.018>
- Inglés, E., Villena, V., Padullés, J. M., Funollet, F., Labrador, V., & Gomila, J. B. (2017). Evaluación Postural Y Análisis Del Equilibrio En Principiantes De Esquí Nórdico. *Revista Internacional de Medicina y Ciencias de la Actividad Física y del Deporte / International Journal of Medicine and Science of Physical Activity and Sport*, 17(68), 651–665. <https://www.redalyc.org/articulo.oa?id=54254647005>
- Joan, x. (2017). *Modifications Wii Balance Board* [Arduino]. <https://github.com/radioelf/Mod.Wii-Balance-Board-> (Obra original publicada en 2017)
- Kozinc, Ž., Löfler, S., Hofer, C., Carraro, U., & Šarabon, N. (2020). Diagnostic Balance Tests for Assessing Risk of Falls and Distinguishing Older Adult Fallers and Non-Fallers: A Systematic Review with Meta-Analysis. *Diagnostics*, 10(9), Article 9. <https://doi.org/10.3390/diagnostics10090667>

- Lima, J. da S., Quadros, D. V. de, Silva, S. L. C. da, Tavares, J. P., & Pai, D. D. (2022). Costs of hospital admission authorizations due to falls among older people in the Brazilian National Health System, Brazil, 2000-2020: A descriptive study. *Epidemiologia e Serviços de Saúde*, 31, e2021603. <https://www.scielo.br/j/ress/a/6Lmf64R4QFSVPLFy8gMJXNq/?lang=en>
- Lorena Cerda, A. (2014). Manejo del trastorno de marcha del adulto mayor. *Revista Médica Clínica Las Condes*, 25(2), 265–275. [https://doi.org/10.1016/S0716-8640\(14\)70037-9](https://doi.org/10.1016/S0716-8640(14)70037-9)
- Martínez, J. G. E. H. (2018). Obesidad en relación con el equilibrio dinámico de mujeres adultas mayores. *Ciencias de la Actividad Física UCM*, 19(2), Article 2. <https://doi.org/10.29035/rcaf.19.2.5>
- Okun, A., Stein, R. E. K., Bauman, L. J., & Silver, E. J. (1996). Content Validity of the Psychiatric Symptom Index, CES-Depression Scale, and State-Trait Anxiety Inventory from the Perspective of DSM-IV. *Psychological Reports*, 79(3), 1059–1069. <https://doi.org/10.2466/pr0.1996.79.3.1059>
- Park, D.-S., & Lee, G. (2014). Validity and reliability of balance assessment software using the Nintendo Wii balance board: Usability and validation. *Journal of Neuroengineering and Rehabilitation*, 11, 99. <https://doi.org/10.1186/1743-0003-11-99>
- Pech-Ciau, B. A., Lima-Martínez, E. A., Espinosa-Cruz, G. A., Pachó-Aguilar, C. R., Huchim-Lara, O., Alejos-Gómez, R. A., Pech-Ciau, B. A., Lima-Martínez, E. A., Espinosa-Cruz, G. A., Pachó-Aguilar, C. R., Huchim-Lara, O., & Alejos-Gómez, R. A. (2021). Fractura de cadera en el adulto mayor: Epidemiología y costos de la atención. *Acta ortopédica mexicana*, 35(4), 341–347. <https://doi.org/10.35366/103314>
- Peydro de Moya, M. F., Baydal Bertomeu, J. M., & Vivas Broseta, M. J. (2005). Evaluación y rehabilitación del equilibrio mediante posturografía. *Rehabilitación*, 39(6), 315–323. [https://doi.org/10.1016/S0048-7120\(05\)74365-6](https://doi.org/10.1016/S0048-7120(05)74365-6)
- Quijoux, F., Nicolaï, A., Chairi, I., Bargiotas, I., Ricard, D., Yelnik, A., Oudre, L., Bertin - Hugault, F., Vidal, P., Vayatis, N., Buffat, S., & Audiffren, J. (2021). A review of center of pressure (COP) variables to quantify standing balance in elderly people: Algorithms and open - access code. *Physiological Reports*, 9(22), e15067. <https://doi.org/10.14814/phy2.15067>
- Rubenstein, L. Z. (2006). Falls in older people: Epidemiology, risk factors and strategies for prevention. *Age and Ageing*, 35(suppl_2), ii37–ii41. <https://doi.org/10.1093/ageing/af1084>
- Santos, D. A., & Duarte, M. (2016). A public data set of human balance evaluations. *PeerJ*, 4, e2648. <https://doi.org/10.7717/peerj.2648>
- Tinetti, M. E., Franklin Williams, T., & Mayewski, R. (1986). Fall risk index for elderly patients based on number of chronic disabilities. *The American Journal of Medicine*, 80(3), 429–434. [https://doi.org/10.1016/0002-9343\(86\)90717-5](https://doi.org/10.1016/0002-9343(86)90717-5)

Sobre los autores

	Azucena Eunice Jiménez Corona. Obtuvo el grado de Doctora en Ciencias en el Instituto de Química, de la Universidad Nacional Autónoma de México. Es Profesora Investigadora de Tiempo Completo en la Universidad Autónoma del Estado de Hidalgo. Sus intereses de investigación incluyen: *Biomarcadores enfocados al diagnóstico por estrés oxidativo. *Estudios en salud pública para la detección de enfermedades crónicas y neurodegenerativas.
	Estefanía Olivares Palacios Realizó sus estudios de Licenciatura en Enfermería por la Universidad Autónoma del Estado de Hidalgo en 2023. Durante todo su proceso educativo participó activamente en la elaboración de artículos científicos en el grupo de investigación biomédica de la Dra. Jiménez-Corona en el desarrollo de investigación en campo clínico.
	Aislinn Joally Ramírez Méndez. Enfermera pasante de la Universidad Autónoma del Estado de Hidalgo. Actualmente es integrante activa del grupo de investigación biomédica de la Dra. Jiménez Corona. Sus intereses de investigación incluyen, el campo clínico en el desarrollo de estudios en poblaciones de adultos mayores.
	Marisa Bautista Lara. Realizó sus estudios de Licenciatura en Enfermería por la Universidad Autónoma del Estado de Hidalgo en agosto del 2023. Participó activamente en el grupo de investigación biomédica de la Dra. Jiménez-Corona en el desarrollo de investigación en campo clínico en el desarrollo de estudios en poblaciones de adultos mayores.
	Zeus Tlaltecutili Domínguez Vega. Obtuvo el título de ingeniero en Mecatrónica por el ITT, el grado de maestro en ciencias por la UAEMex y recibo el grado de Ph.D en ciencias biomédicas por la Universidad de Groningen en Países Bajos en 2022. Actualmente es profesor investigador posdoctorante en la UAEMex. Sus intereses incluyen análisis y procesamiento de datos biomédicos y el uso tecnologías para la mejora en la calidad de vida.
	Rigoberto Martínez Méndez obtuvo el título de ingeniero en electrónica, el grado de maestro en ciencias en bioelectrónica por el CINVESTAV, México y el grado de Ph.D en biomédica por la Universidad de Chiba. Actualmente es miembro del SNI nivel I y se desempeña como profesor e investigador en la UAEMex. Sus intereses de investigación incluyen el análisis de bioseñales y el desarrollo de dispositivos biomédicos portables no invasivos.



Recibido 26 Ene. 2023

ReCIBE, Año 12 No. 1, May. 2023

Aceptado 09 Sep. 2023

Fundamentos generales para la Protección de Software y Programas de Cómputo en México

General Fundamentals for the Protection of Software and Computer Programs in Mexico

Alfredo Aguilar Ruiz¹
alfredsbox@hotmail.com

¹Universidad de Guadalajara

Resumen: *La siguiente investigación tiene como objetivo el dar a conocer las diversas alternativas que existen en México para la protección de software y programas informáticos. Esto con la finalidad de proteger legal y comercialmente cualquier herramienta desarrollada dentro de las pequeñas y medianas empresas que sirva para la operatividad y funcionamiento de estas. Además de que la mayoría de las veces en estas herramientas se plasma el know how de la compañía, e incluso puede representar o contribuir a crear la ventaja competitiva la empresa.*

En esta investigación de tipo cualitativa y descriptiva se presentarán las diversas opciones que se tienen en México para la protección y blindaje de programas informáticos, así como las implicaciones y requisitos que debe cumplir la obra para poder estar registrada bajo los distintos esquemas. Asimismo, se darán a conocer las instancias internacionales que también deben contemplarse para brindar una protección integral a los softwares y programas informáticos.

Palabras Clave: *Registro de software, derechos de autor, patentes de software, protección de softwares, propiedad intelectual, autoría de un software.*

Abstract: *This document is descriptive qualitative research that presents the different options in Mexico for legal protection of software and computer programs, as well as the implications and requirements that the program must meet in order to be registered under the different schemes. Likewise, international instances are suggested that should also be considered to fully protect the software.*

Keywords: *Software registration, copyright, software patents, software protection, intellectual property, software authorship.*

1. Introducción

Es bien sabido que el departamento de tecnologías de la información (TI) en una empresa, suele ser el área que brinda estructura y soporte a toda la organización y la que hace posible que la gran mayoría de los procesos interdepartamentales se desarrollen en óptimas condiciones, automatizando los procesos clave que generan valor a los usuarios y empresa por igual. Para dar cumplimiento a esta ardua labor, se debe contar con sistemas informáticos que puedan sostener estas operaciones, es decir, software y toda la parte blanda que es inherente al área de tecnologías de la información y comunicaciones. Muchas empresas se apoyan con terceros para el desarrollo de software, ya que prefieren dejar a los expertos esta labor; otras optan por utilizar softwares comerciales a través del arrendamiento o adquisición de licencias por medio de *partners* que permiten usar dichas plataformas por tiempo determinado, y muchas otras prefieren desarrollar estos programas *inhouse* con el fin de minimizar costos y de preservar información de carácter confidencial. Bajo el escenario en donde las PyMEs apuestan por desarrollar su propio sistema (ya sea con un externo o *inhouse*), es conveniente que conozcan las implicaciones legales en materia de propiedad intelectual y sobre los derechos de autor que deben contemplar para proteger su sistema, ya que la mayoría de organizaciones suelen obviar esta parte, lo que se traduce en que los prestadores de servicio se atribuyan el 100 por ciento del desarrollo como propio y que incluso no compartan el código fuente. Es aquí donde radica la importancia de conocer cuáles son los derechos y las distintas opciones que se tienen para proteger un software de ser plagiado y de que sea utilizado con fines de lucro por un tercero.

Es importante aclarar que la finalidad de este documento no es debatir sobre los derechos de autor en un software, ni sobre los derechos y obligaciones de un desarrollador y la empresa que contrató el servicio. El objetivo de esta investigación es dar a conocer las diferentes alternativas que se tienen para la protección legal de software en el país y conocer sus requisitos, costos, entidades responsables y pormenores sobre cada uno de los trámites de manera puntual; también se presentan algunos escenarios a considerar para optar o no por la solicitud de patente de software o programa de cómputo, que contribuya a dar claridad sobre este tema del cual existe muy poca información a nivel México. De igual forma, megatendencias como el *cloud computing*, el *software as a service (SaaS)*, el *Internet of Things (IoT)*, el *Big Data* y todos los productos derivados y generados a través de estas tecnologías exponenciales, representan un reto a nivel jurídico que debe ofrecer cobertura y protección al mismo ritmo vertiginoso en que han evolucionado los sistemas.

Además, esta investigación es pertinente en virtud de los más de tres millones de solicitudes de patente que se realizan cada año de acuerdo con cifras publicadas por la Organización Mundial de la Propiedad Intelectual (2018), en donde el rubro que más crecimiento ha alcanzado en los últimos años es el relacionado a tecnología informática en donde se sitúan las patentes de software; lo cual representa una enorme área de oportunidad para las empresas mexicanas que están desarrollando programas de cómputo con la intención de comercializarlo a nivel nacional e internacional.

2. Desarrollo

Primeramente, se tiene que definir el concepto de software *per se*, que de acuerdo con la Real Academia Española (2020), es un anglicismo empleado en la informática que significa “conjunto de programas, instrucciones y reglas informáticas para ejecutar ciertas tareas en una computadora”. También puede referirse a un listado de programas escritos en un lenguaje de programación para implementar un algoritmo, o al código binario cargado en un aparato basado en una computadora, y también puede abarcar la documentación adjunta (Ruiz Salazar, 2016).

Vale la pena comenzar este artículo señalando que propio el carácter técnico de los softwares y programas informáticos propició que tradicionalmente se defendiera su carácter industrial. Sin embargo, tras el Convenio de Berna se estableció que la figura jurídica responsable de proteger el software sería la de propiedad intelectual. Dicho lo anterior y de acuerdo con las leyes internacionales, un software o programa informático como tal no es una invención patentable. Las patentes no se conceden sólo por listas de programas; los listados de programas como tales están protegidos por derechos de autor.

Por su parte, el artículo 19 de la *Ley de la Propiedad Industrial en México*, excluye la patentabilidad de los programas de computación (Cámara de Diputados, 2018), mientras que el artículo 13 de la *Ley Federal de Derechos de Autor* reconoce el derecho de autor de los programas de cómputo y bases de datos en el inciso XI (Instituto Nacional del Derecho de Autor, 2018).

2.1 Antecedentes del registro del Software en México

Es muy importante recordar que la *Ley Federal del Derecho de Autor* publicada desde el año de 1996, en su artículo 101 se encarga de regular a los programas de cómputo con una protección igual a la que se otorga a las obras literarias o artísticas (INDAUTOR, 2018). Esto como consecuencia de que México se adhirió a la *Organización Mundial de Comercio* el primero de enero del año 1995, lo cual implicó sujetarse al anexo de “Aspectos de Derechos de Propiedad Intelectual Relacionados con el Comercio”. En este acuerdo se estableció que los programas de cómputo serían protegidos en la misma medida que las obras literarias por virtud del Convenio de Berna de 1971.

Muchos autores y expertos en propiedad intelectual argumentan que al final del día el software es un lenguaje literario, al tratarse de un algoritmo matemático que da instrucciones a una computadora para que ejecute una acción, por lo que a ese lenguaje de programación se le debe considerar como literatura. En contraste, las creaciones u obras literarias tienen entre sus principales características el reflejar la personalidad de su autor, mientras que un software resuelve un problema técnico a través de un código fuente en el campo de la informática, que suele poseer una gran carga de actividad inventiva, requisito indispensable para patentar una invención (Hernández, 2016). Aunque la *Ley de la Propiedad Industrial* excluye la patentabilidad de los programas de computación, es más que evidente que un software especializado puede reunir los tres requisitos de patentabilidad como son la novedad, la actividad inventiva y la aplicación industrial.

En contraste, Estados Unidos que también forma parte de la *Organización Mundial de Comercio* desde 1995, sí otorga patentes sobre software. Por su parte, la Oficina Europea de Patentes y Marcas también otorga patentes a invenciones implementadas en un ordenador cuando estas resuelven un problema técnico, es decir, en aquellos casos en los que el programa gestiona un proceso industrial o el funcionamiento de una máquina (Hernández, 2016).

Es importante que los creadores de software y las empresas en general tengan claridad sobre estas diferencias, ya que a nivel México no se conceden patentes de software sin importar que Estados Unidos ya exista una. Esto suele causar confusión y dolores de cabeza a los fabricantes y desarrolladores de software, ya que asumen que la protección de patente en otros países hará posible su otorgamiento en México, como ocurre generalmente con otros tipos de patentes (Isla y Villaseñor, 2020).

2.2 Registro de Software: Derecho de Autor

En México las obras literarias como cuentos, novelas, poemas, películas, pinturas, obras de teatro, programas de cómputo, entre otros, son resultado de la capacidad creativa de sus autores, cuyos derechos son protegidos a través de los derechos de autor. Estos derechos confieren a los autores el disponer libremente de sus obras, lo que se traduce en que pueden decidir su mantenerlas inéditas, publicarlas, heredarlas, comercializarlas o incluso transferir derechos a terceros para que comercialicen con ellas (Justia México, 2020).

La comercialización de obras artísticas y literarias producen beneficios económicos para los autores, para todos los involucrados en las industrias y por consiguiente también beneficio económico para los países. En México, la *Ley Federal del Derecho de Autor* protege los derechos de los autores con relación a sus obras otorgándoles derechos de carácter personal y patrimonial. Es importante destacar que la ley protege las obras desde el momento en que las mismas se plasman en un soporte material, y el reconocimiento de los derechos de autor no requiere de registro, documento o formalidad alguna.

En el caso de software o programa de cómputo, la protección del derecho de autor es automática desde el momento mismo en el cual se termina la creación del programa, o de cada uno de sus respectivos componentes o elementos; incluso antes de compilar el código fuente, el programa ya se encuentra protegido por el Derecho de Autor (INDECOPI, 2013).

2.2.1 Características de la Obra

De acuerdo con el artículo tercero de la *LFDA* publicada por INDAUTOR (2018) sobre las características de las obras protegidas por esta Ley, y a los requisitos que debe cumplir un software de acuerdo a lo publicado por INDECOPI (2013), a continuación, se mencionan los atributos que debe presentar el software a registrar:

a) *Ser susceptible de ser divulgado*

En términos de software se refiere a que pueda ser “leído” en su propio lenguaje de programación o mediante el medio para el cual el programa haya sido creado.

b) *Ser original*

El software debe ser el resultado de un proceso creativo y su código fuente debe ser capaz de diferenciarse de cualquier otro programa de cómputo que haya sido creado con anterioridad, aun cuando hayan sido creados con el mismo lenguaje de programación, con las mismas herramientas o estén orientados a realizar los mismos procesos o a obtener los mismos resultados.

c) *Ser susceptible de ser reproducido*

Se refiere a que dicho programa pueda grabarse o fijarse en algún soporte que permita obtener copias de este, sin perjuicio de que contenga impedimentos o restricciones tecnológicas que restrinjan dicho acto.

2.2.2 Derechos morales y patrimoniales

Los derechos morales son aquellos que corresponden al autor sobre la obra de su creación, y permiten al autor decidir si su obra ha de ser divulgada o mantenerse inédita, y en caso de ser divulgada de qué forma se realizará, ya sea como una obra anónima o bajo un seudónimo. Además, los derechos morales permiten al autor exigir el reconocimiento de su calidad de autor respecto de la obra por él creada, modificar la obra, oponerse a cualquier modificación o alteración a la misma, e incluso retirar su obra del mercado. Es importante destacar que el autor es el único, primigenio y perpetuo titular de los derechos morales sobre su obra; derechos que son inalienables, imprescriptibles, irrenunciables e inembargables al autor (Justia México, 2020).

Por otro lado, los derechos patrimoniales hacen referencia a la capacidad que tiene un autor para explotar su obra de manera exclusiva o de autorizar a terceros su explotación. En cuanto a la titularidad de los derechos patrimoniales, la poseen el autor de la obra, sus herederos o adquirentes por cualquier título, es decir, por virtud de algún acto o negocio jurídico. Los derechos de explotación de una obra pueden transmitirse por medio de actos, contratos, convenios o licencias de uso, donde se establezca el monto de la remuneración, el procedimiento para fijarla y los términos para su pago (Justia México, 2020).

Una vez comprendido la diferencia entre derechos morales y patrimoniales, y a quién o quiénes pertenece cada uno, en el marco de una empresa que desarrolla un software y que le interesa registrarlo, la *LFDA* dispone que si una obra es creada como consecuencia de una relación laboral en la cual existe un contrato individual de trabajo por escrito, se presumirá que los derechos patrimoniales se dividen por partes iguales entre el patrón y él o los trabajadores que crearon el software. Por su parte, el patrón podrá divulgar la obra sin autorización del trabajador, pero el trabajador no podrá divulgar la obra sin autorización del patrón. En el escenario que no existiera contrato individual de trabajo por escrito, los derechos patrimoniales de la obra corresponden al trabajador (Justia México, 2020).

Como dato adicional, cabe resaltar que la vigencia de los derechos patrimoniales es vitalicia (durante toda la vida del autor), y cien años más tras la muerte del autor; cuando la obra pertenezca a varios coautores, los cien años se contarán a partir de la muerte del último coautor. Al concluir la vigencia de los derechos patrimoniales la obra pasará al dominio público (Justia México, 2020).

2.2.3 Requisitos para registrar un software

Aunque no es obligatorio el registro de una obra de estas características ante el INDAUTOR, es altamente recomendable su registro, ya que otorga protección jurídica a los autores y titulares de derechos frente a posibles actividades indeseadas de terceros. De acuerdo con lo publicado en el portal de internet de INDAUTOR (2020) con relación al registro y a lo que se considera un programa de cómputo, se refiere a la expresión original en cualquier lenguaje o código, de un conjunto de instrucciones que con una estructura y organización determinada logran que una computadora o dispositivo realice una función específica.

La solicitud de registro de la obra debe realizarse ante la *Dirección del Registro Público del Derecho de Autor* que forma parte del INDAUTOR. Dicho trámite puede ser realizado por el autor de la obra, por el titular de derechos, o bien el representante legal (a través de instrumento notarial) o gestor (mediante carta poder).

Los requisitos para registrar un software son los siguientes (INDAUTOR, 2020):

- Copia del programa de cómputo (archivo ejecutable) en soporte electrónico (CD o DVD)
- Breve descripción del programa, donde se presenten cuáles son sus funciones, para qué sirve, tecnologías implementadas, lenguajes de programación, etc.
- Título de la obra (para efectos de identificar frente a otros programas).
- Mínimo 10 líneas de código fuente (5 primeras líneas y 5 últimas).
- Llenar formato RPDA01, y en su caso el auxiliar (cuando hay coautores), por duplicado.
- Si la obra es por encargo, el documento que acredite esta situación (P.Ej.: contrato de obra, de prestación de servicios, de colaboración, de cesión, etc.).
- Comprobante de pago de derechos por \$271 pesos.

2.2.4 Sobre la autoría del Software

Aunque a nivel México existe muy poca información sobre los detalles de derecho de autor de software o programas de cómputo, en específico sobre la autoría y sobre quién o quiénes deben considerarse autores, en virtud de que los programas suelen desarrollarse por encargo de un tercero y creado por un equipo de trabajo que muchas veces permanece en el anonimato.

Esta falta de información y especificación en materia de derecho de autor de software está directamente relacionada con la falta de producción de programas de cómputo en el país, que de acuerdo con Lizardi (citado en Hernández, 2001), en México se desarrolla muy poco software por falta de recurso humano calificado, por el costo que representa ya que los buenos programadores son escasos y caros, por desconocimiento comercial y por la piratería que existe que no permite competir. No obstante, y para fines de contemplar los diferentes escenarios y actores a considerar en materia de autoría de software, se tomó información publicada por el Instituto Nacional de Defensa de la Competencia y de la Protección de la Protección Intelectual de Perú, en donde detallan a profundidad a quién o quiénes reconocer como autores en materia de software.

De acuerdo con lo publicado en la Guía de Derecho de autor para creadores de software del INDECOP (2013), se deberá reconocer como autor de un programa informático o software a toda aquella persona que haya participado en su creación o programación, independientemente de si es un ingeniero de sistemas con estudios especializados, un programador o una persona con conocimientos básicos. A su vez, podrán ser considerados como co-autores, todas aquellas personas que escribieron o programaron directamente, ya sea todo o en parte, el código fuente del software en cuestión; así como todas aquellas personas que programaron o diseñaron la parte gráfica del mismo. Incluso si el programa de ordenador en cuestión fue desarrollado por varias personas trabajando en forma separada y sin conocerse, todas ellas también deberán ser consideradas como co-autores de tal obra (INDECOP, 2013).

Cabe mencionar que en México y de acuerdo con lo publicado por INDAUTOR (2018), las obras se registran y protegen también por el número de creadores que intervienen, pudiendo clasificarse por:

- a) *Individuales*
Se refiere a las obras que fueron creadas por una sola persona.
- b) *De colaboración*
Obras que han sido creadas por varios autores.
- c) *Colectivas*
Se refiere a obras creadas por la iniciativa de una persona física o moral que las publica y divulga bajo su nombre y dirección, en las cuales la contribución personal de los diversos autores que han participado en su elaboración se funde en el conjunto con vistas al cual ha sido concebida, sin que sea posible atribuir a cada uno de ellos un derecho distinto e indiviso sobre el conjunto realizado.

2.3 Patente de Software: Propiedad Industrial

Una patente es el derecho que el gobierno de un país le otorga a un inventor de impedir a otras personas la venta, fabricación o uso en dicho país del invento en cuestión por un período de tiempo determinado. Si se opta por no explotar la patente, esta se puede vender o ceder los derechos a un tercero para que éste último la comercialice bajo licencia (Instituto Nacional de Propiedad Industrial, 2020).

A pesar de que de acuerdo con la Ley de Propiedad Intelectual en su artículo 19 expresamente refiere que los programas de computación no serán considerados invenciones y por lo tanto no son patentables en México, es posible integrar un software a una patente. De acuerdo con Zárate (2019) para lograr este objetivo es necesario que el software o programa informático esté expresado como método o proceso, ya que la protección otorgada a través de la patente será propiamente al método o proceso y no al software como tal, al cual se le considera como el conjunto de instrucciones plasmadas en forma de código o lenguaje de programación. En otras palabras, la protección de la patente al método o proceso abarcaría la funcionalidad del software, pero no al código o forma de expresión, ya que esto es considerado una obra de acuerdo con Derecho de Autor.

Es importante tener claro que será posible integrar el software a la protección de la patente siempre que éste sea incorporado en alguna de las reivindicaciones que se especifican en la solicitud de patente, donde deberá describirse el papel o relevancia del software en la invención. Además, obtener esta protección será posible siempre y cuando el software forme parte integral de la invención, y que sin ésta dicha invención no pudiera funcionar.

Como referencia y para ejemplificar este escenario, se pueden considerar los sistemas embebidos, es decir, subsistemas que se encuentran inmersos en un dispositivo más grande, que operan de manera independiente y que cubren necesidades específicas. Estos sistemas son diseñados para realizar pocas funciones en tiempo real y se les puede encontrar en los autos, teléfonos celulares e incluso en electrodomésticos comunes como refrigeradores y microondas.

2.3.1 Requisitos para solicitar una patente

En México, la autoridad competente en lo que a patentes y demás figuras de Propiedad Industrial se refiere es el Instituto Mexicano de la Propiedad Industrial. A continuación, se mencionan los requisitos formales para la solicitud de una patente en México (Gobierno de México, 2020):

- Nombre y domicilio del inventor.
- Nombre, domicilio y nacionalidad del solicitante.
- Fecha de divulgación previa (opcional).
- Datos del representante o apoderado legal (opcional).
- Denominación de la invención
- Comprobante de pago por presentación de solicitud de patente y sus anexos de \$4550 pesos
- Llenar formato IMPI-00-009 denominado “Solicitud de Patente de Invención o de Registro de Modelo de Utilidad o de Registro de Diseño Industrial”.
- Anexar descripción clara y completa de la invención.
- Dibujos para comprender la descripción.
- Una o más reivindicaciones claras y concisas, que no podrán exceder del contenido de la descripción.
- Resumen de la descripción de la invención, que servirá únicamente para su publicación y como elemento de información técnica.
- Traducción de los documentos presentados en idioma distinto al español

En caso de otorgarse la patente, esta tendrá una vigencia de 20 años improrrogables contados a partir de la fecha de presentación de la solicitud.

2.3.2 Alternativas para la protección de Software: Patentar en Estados Unidos

A pesar de que existen tratados internacionales sobre Propiedad Intelectual, cada país puede otorgar protección legal pero solo dentro de las fronteras de dicho país. Eso significa que, si un invento ha sido patentado en México, Colombia y Brasil, un tercero puede solicitar la patente en Estados Unidos y producirlo, venderlo y generar ingresos sin tener que pagar o reconocer al inventor. Para tener el derecho de exclusividad de fabricación, venta y producción de un invento, así como para comercializar el licenciamiento de este, se debe contar con una patente en cada país en el que desee esta protección.

En el caso de los Estados Unidos, el gobierno federal garantiza este derecho a través de un conjunto de leyes y reglas amparadas por la Constitución, que facultan al Congreso para fomentar el progreso de la ciencia y las artes útiles, asegurando a los autores e inventores, por un tiempo limitado, el derecho exclusivo sobre sus respectivos escritos y descubrimientos; la Oficina de Patentes y Marcas de los Estados Unidos (USPTO), dependiente del Departamento de Comercio, es el organismo responsable de la solicitud y expedición de patentes (Patentes USA, 2020).

Esto se traduce en que el Gobierno le otorga a un inventor el derecho a una patente a cambio de que éste revele todos los detalles acerca de su invento, desde cuáles son sus componentes y cómo se fabrica o produce, hasta cómo utilizarlo. De esta manera, pasado el término de la patente, toda la sociedad se ve beneficiada por el invento. La patente da a su titular el poder de impedir o evitar que otras personas puedan comercializar su invento, salvo que cuenten con una licencia o cesión por parte del titular. Este derecho dura entre 15 y 20 años, dependiendo del tipo de patente, tras los cuales el invento pasa al dominio público (Patentes USA, 2020).

Cabe destacar que una patente por sí misma es una de las mejores maneras de obtener ventaja competitiva en cualquier mercado, ya que otros se ven impedidos de copiar la idea, a la vez que permite recuperar la inversión que se ha realizado en crear el invento. Debido a esto, las patentes incentivan la innovación a todos los niveles, ya que promueven el desarrollo de nuevos inventos al garantizar una protección legal que permitirá recuperar la inversión en tiempo, dinero y esfuerzo, a la par de obtener una ganancia económica.

2.3.2.1 Consideraciones para patentar en Estados Unidos

Debido en gran medida a su nivel de innovación, influencia a nivel global y fortaleza económica, Estados Unidos es el país en el que la mayoría de las empresas e inventores suelen patentar sus inventos y creaciones, al representar el país y mercado que mayor impacto económico puede generar (Patentes USA, 2020).

En el caso de inventores o empresas extranjeras que deseen patentar en los Estados Unidos, se puede realizar la solicitud mediante la representación de cualquier Agente de Patentes o Abogado de Patentes registrado ante la USPTO; en el escenario de que el inventor/creador haya cedido o tenga obligación de ceder sus derechos sobre el invento a una empresa (P.Ej.: si el inventor era empleado de dicha empresa y el invento se generó en el marco del contrato laboral) tampoco es requisito que dicha empresa sea de origen estadounidense. De hecho, un poco más del 30% de las patentes que se conceden en los Estados Unidos corresponden a solicitantes extranjeros.

En cuanto a los costos por obtener una patente en Estado Unidos, oscilan entre los 4500 y los 7000 dólares, tomando en consideración el tipo de patente solicitada, las tarifas asociadas a la búsqueda, examinación y concesión de la misma, al número de reivindicaciones, a las extensiones en los plazos de respuesta a las acciones oficiales, cuyas tasas oficiales están publicadas en el tarifario de la USPTO (s.f.); asimismo, se consideran los honorarios por el asesoramiento legal de un agente o representante autorizado ante la USPTO.

2.3.3 Controversias sobre Patentes de Software

Hoy día existe una gran polémica sobre qué alcance debe concederse a las patentes de software, mejor conocidas como “invenciones implementadas por computadora”. De acuerdo con los detractores de las patentes para software, todos los programas informáticos se componen de millones de componentes, entre procedimientos, métodos y algoritmos, los cuales podrían ser patentables por sí mismos, o incluso estar ya patentados en otras invenciones (Isla y Villaseñor, 2020). A su vez, resultaría prácticamente imposible establecer si un código de programación infringe alguna patente, ya que para tener certeza de esto se requeriría analizar y evaluar de manera puntual todas las patentes de software que existen, y aun así sería difícil comprobar dicha infracción. De igual forma, se requeriría de un proceso judicial para poder determinar con certeza y veracidad si una patente está siendo infringida. Por el gran esfuerzo humano y económico que lo anterior representa, la búsqueda exhaustiva y un pleito de patentes se encuentra fuera del alcance de las PyMEs.

Por otro lado, los desarrolladores que apoyan el uso de “software libre” argumentan que el empleo de patentes en software impediría el desarrollo de muchos proyectos en los que no se paga licencia. Desde el punto de vista social, la implementación de patentes en software privatizaría el conocimiento al excluir a la mayoría de la población como productores e incluso como consumidores de los objetos de dichas patentes (Isla y Villaseñor, 2020).

2.3.4 Medidas adicionales para la Protección de Software

De acuerdo con Nadal (2020), antes de lanzar un software al mercado o *“liberarlo en productivo”* como se le conocer en el argot e los desarrolladores, es conveniente tomar ciertas medidas de seguridad para proteger el mismo, entre las que destacan las siguientes:

1. Controlar que sólo unas cuantas personas de mucha confianza tengan acceso al código fuente; de esta manera el programa de software podrá considerarse como secreto empresarial.
2. Generar contratos de confidencialidad que deberán firmar todas las personas que trabajen con el software o en el desarrollo, independientemente si son empleados de la empresa o si son externos, en donde se comprometan a no compartir la información ni los secretos que se manejan, así como reconocer que el software es propiedad de la empresa.
3. Es conveniente que el software esté alojado en un servidor que sea propiedad de la empresa; esto con el fin de prevenir que se tenga acceso a este código y que un *hoster* pueda desarrollar un servicio o un programa similar.
4. Adicional, se recomienda el registro de la identidad y nombre del software a través del registro de marca, que contempla el registro del nombre del software, de su logotipo y de los colores utilizados en su imagen; el registro de marca en México se realiza ante el IMPI.

3. Discusión

Esta investigación se abordó bajo un enfoque cualitativo de tipo descriptivo, ya que no buscó medir, sino cualificar y describir el fenómeno a partir de sus rasgos determinantes (Bernal, 2006); a su vez, buscó especificar las propiedades y características de un proceso y revelar nuevas interrogantes en el proceso de interpretación (Hernández, Fernández y Baptista, 2014).

La metodología empleada fue la revisión a la literatura en materia de derecho de autor y patentes para software en México y en Estados Unidos. A continuación, se presentan algunos consejos sugeridos por la Organización Mundial de la Propiedad Intelectual (2020), que conviene tomar en cuenta al considerar la protección por medio de patentes a los programas de cómputo:

1. *Analizar si realmente se necesita una patente para una invención asociada a una aplicación o un sistema informáticos.*

En muchos países los programas informáticos están protegidos por el derecho de autor y no se requiere ninguna formalidad como el registro o el depósito de ejemplares en los 151 países que forman parte del Convenio de Berna para la Protección de las Obras Literarias y Artísticas, lo cual les confiere una ventaja por su sencillez. Asimismo, el autor goza de un período de protección que dura toda su vida más otros 100 años tras haber fallecido el autor.

En contraste, las solicitudes de patente deben presentarse para cada país en el que se desee obtener protección y dicha solicitud debe cumplir una serie de requisitos que pueden ser jurídica y técnicamente complejos. La duración de la protección por medio de patentes suele ser de 20 años a partir de la fecha de presentación de la solicitud, es decir, mucho más corta si se compara con la protección del derecho de autor.

A menos que los recursos financieros no representen una limitante, vale la pena considerar si patentar una innovación asociada a un software es la mejor manera de proteger el producto, y valorar la posibilidad de utilizar otros tipos de propiedad intelectual, como los dibujos o modelos industriales, la protección del secreto comercial y el registro de marca.

2. *Identificar la parte esencial de la innovación y determinar qué es lo que se desea proteger contra los competidores*

Los sistemas o programas de cómputo pueden formar parte de un ordenador u otro tipo de aparato y con frecuencia ese software se desarrolla, reproduce y distribuye a través de otro tipo de soportes internos o externos. Estos programas realizan funciones como controlar un dispositivo o como interfaces entre un ordenador y un ser humano; también sirven para procesar datos, e incluso pueden emplearse para explorar nuevas teorías científicas.

Una vez comprendido el uso y la relación del software con el hardware es posible tener mayor claridad sobre lo que se desea proteger de los competidores; es decir, determinar cuál es la parte que aporta mayor valor en una innovación asociada a un programa de cómputo para valorar las posibilidades de obtener o no una patente, y es que este valor puede hallarse en el dispositivo *per se*, en el sistema, en un algoritmo, en un método, en el procesamiento de los datos o en el propio software.

3. *Validar si la innovación es patentable, ya que no todos los programas de cómputo pueden ser protegidos con una patente.*

Para poder patentarse, una invención debe cumplir por lo menos cinco criterios, los cuales determinan su patentabilidad:

- a) La invención debe constituirse por una materia patentable.
- b) La invención debe ser susceptible de aplicación industrial.
- c) La invención debe ser nueva (original).
- d) Debe implicar una actividad inventiva, es decir, no ser evidente.
- e) La divulgación de la invención en la solicitud de patente debe cumplir una serie de normas de forma y de fondo.

De acuerdo a los criterios mencionados, no es suficiente con que la invención sea nueva, es decir, que sea diferente de lo que exista hasta ese momento en el estado de la técnica, sino que la diferencia debe ser significativa y esencial a la invención. Tras lo anterior, es necesario considerar que no será posible obtener una patente para un software que solo sustituya las soluciones técnicas y físicas existentes por las mismas soluciones empleando código y una computadora, si esta sustitución resulta evidente para cualquier ingeniero especializado en ese campo técnico.

4. *Evaluar si se requiere protección en el extranjero, ya que los requisitos para obtener una patente suelen diferir de un país a otro.*

Si se desea patentar una innovación en el extranjero, es necesario conseguir una patente en cada país donde interese proteger la invención; es decir, una patente concedida en cierto país sólo tendrá validez en ese territorio y no es posible impedir un competidor utilice dicha invención en otros países.

En caso de requerir patentar una innovación en varios países, y con el fin de simplificar el procedimiento y reducir costos, el *Tratado de Cooperación* en materia de Patentes establece una opción para facilitar la presentación de solicitudes internacionales, en la cual un solicitante puede presentar una única solicitud internacional que tiene el mismo efecto que las solicitudes nacionales presentadas en cada uno de los Estados del Tratado de Cooperación en materia de Patentes (PCT). Sin embargo, se tiene que considerar que, aunque la solicitud será un trámite único común, depende de cada Estado el que la patente sea concedida para cada territorio.

5. Consultar a un experto en materia de propiedad intelectual

Cumplimentar adecuadamente una solicitud de patente es un reto para un experto en propiedad intelectual que conozca la legislación y la práctica nacional. Es fundamental redactar correctamente la solicitud de patente desde el inicio, ya que una vez presentada, las posibilidades de corregirla serán mínimas. Sobre todo, es esencial redactar bien las peticiones o "reivindicaciones" de la solicitud, que definen el objeto o los objetos cuya protección se está solicitando.

Cabe resaltar que una vez que se concede la patente, el cumplimiento puntual de los derechos que confiere se relacionará con la definición escrita de la invención que se haya presentado en las reivindicaciones. Por tal motivo, se recomienda ampliamente consultar con un experto en propiedad intelectual para la redacción de la solicitud de patente; de manera específica, en las innovaciones asociadas a programas de cómputo, si las reivindicaciones están bien redactadas se evitará la denegación de la solicitud.

Bajo este tenor, es importante contemplar que en algunos países se aceptan peticiones en las que se define un programa informático por sí solo o registrado en un soporte legible por máquina (como un disco o CD-ROM). Esto debido a que frecuentemente los softwares se comercializan en algún medio legible por una computadora o a través de Internet, de manera independiente del hardware. Para impedir la comercialización no autorizada de estos productos, algunos países permiten reclamar la propiedad de un programa por sí solo o almacenado en algún medio legible por máquina. En países donde no se aceptan estos reclamos de propiedad, como el caso de México, se tienen que redactar cuidadosamente las reivindicaciones de la solicitud y precisar si es en forma de un aparato, un proceso o un sistema, que abarque las ideas funcionales subyacentes al software y que tome la parte esencial de la invención.

4. Conclusiones

El objetivo central de esta investigación fue presentar las diferentes alternativas que se tienen para la protección legal de software a nivel nacional y detallar sus requisitos, costos, instancias responsables y pormenores sobre cada uno de los tipos de trámites que se pueden realizar, que ayude a las PyMEs a obtener protección legal de un software o programa de cómputo. También se detallaron los atributos que debe poseer dicho software para ser protegido y los pormenores en materia de derechos de autor; así como algunos escenarios a considerar para optar o no por la solicitud de patente de software.

De acuerdo con la información presentada en esta investigación es posible concluir lo siguiente:

Un software desde que se realiza cuenta con protección de derechos de autor, misma que tiene efectos legales, aunque la obra no se encuentre registrada formalmente ante el INDAUTOR. En caso de requerirse una patente para proteger el software como una invención, ya sea porque se esté interesado en explotarlo comercialmente a nivel nacional e internacional o ceder el licenciamiento a un tercero, se cuentan con las siguientes alternativas para solicitar su patente en México:

- *Se podrá solicitar la patente ante el IMPI, siempre y cuando el software que se desea proteger se presenta como un sistema embebido, es decir, que sea parte de un aparato un dispositivo cuyo funcionamiento dependa del software.*
- *También se puede optar por solicitar la patente como método, modelo, o proceso que se realiza a través de un software.*
- *A nivel internacional, y en caso de que se requiera por estrategia comercial, se recomienda realizar la solicitud de patente en los Estados Unidos, a fin de que el sistema cuente con una protección legal fuera del país.*

También, se sugiere tomar en cuenta las recomendaciones planteadas por la OMPI sobre patentabilidad de un software, entre las que destacan: *validar si realmente se requiere una protección a través de una patente, evaluar si se cuentan con los recursos de tiempo y dinero suficientes para realizar el trámite, validar si el software cumple con todos los criterios para ser patentable, determinar cuál es la innovación que realmente se quiera proteger, ya que esto será fundamental para conocer se requiere o no una patente de software, y asesorarse con un experto en materia de propiedad industrial del país donde se desee tramitar la patente.*

5. Referencias

Bernal, C. A. (2006). *Metodología de la Investigación*. Bogotá, Colombia: Pearson. Cámara de Diputados

(2018). *Ley de Propiedad Industrial*.

http://www.diputados.gob.mx/LeyesBiblio/pdf/50_180518.pdf

Gobierno de México (2020). *Solicitud de Patente de Invención o de Registro de Modelo de Utilidad o de Registro de Diseño Industrial*. Recuperado de:

https://www.gob.mx/cms/uploads/attachment/file/493901/IMPI_00_009_2019_editable.pdf

Hernández, E. (2016, 25 de mayo). ¿Por qué no podemos patentar el software? Razones técnicas y económicas.

Revista Digital sobre Patentes, Marcas y Propiedad Intelectual.

Recuperado de <https://www.mipatente.com/por-que-no-podemos-patentar-el-software-razones-tecnicas-y-economicas/>

Hernández, J. (23 de diciembre de 2001). ¿Desarrollas software? ¡Regístralo! *Reforma*.

Recuperado de: <https://glib.org.mx/article.php?story=20011226135147472> Hernández, R.,

Fernández, C., y Baptista, P. (2014). *Metodología de la Investigación*. D.F.,

México: McGraw-Hill.

INDECOPÍ (2013). *Guía de Derecho de Autor para Creadores de Software*. Recuperado de:

https://www.indecopi.gob.pe/documents/20182/143803/GDA_CreadoresDeSoftware.pdf

Instituto Nacional del Derecho de Autor (2018). *Ley Federal del Derecho de Autor*.

<https://www.indautor.gob.mx/documentos/marco-juridico/leyfederal.pdf>

Instituto Nacional del Derecho de Autor (2020). *Registro de obra programa de cómputo*.

Recuperado de: https://www.indautor.gob.mx/tramites-y-requisitos/registro/obra_computo.html

Instituto Nacional de Propiedad Industrial (2020). ¿Qué son las patentes? Recuperado de:

<https://www.inapi.cl/portal/institucional/600/w3-article-744.html>

Isla, G. y Villaseñor, C. (2020). Patentes de Software. Conociendo la Legislación. *Software*

Guru, 17. Recuperado de: <https://sg.com.mx/revista/17/patentes-software-conociendo-la-Legislación>

Justia México (2020). *Preguntas y Respuestas Sobre Derechos de Autor*. Recuperado de:

<https://mexico.justia.com/derecho-de-la-propiedad-intelectual/derechos-de-autor/preguntas-y-respuestas-sobre-derechos-de-autor/#q31>

Nadal, S. (2020, 04 de mayo). *¿Software y Propiedad intelectual?* Recuperado el 5 de mayo de 2020 de <https://www.mipatente.com/software-y-propiedad-Intelectual/>

Organización Mundial de la Propiedad Intelectual (2018). *Datos y cifras de la OMPI sobre PI, edición de 2018*.

Recuperado de: https://www.wipo.int/edocs/pubdocs/es/wipo_pub_943_2018.pdf

Organización Mundial de la Propiedad Intelectual (2020). *Patentar los programas informáticos*.

https://www.wipo.int/sme/es/documents/software_patents.htm

Patentes USA (2020). *La importancia de patentar sus inventos en los Estados Unidos de América*. Recuperado de:

<https://www.patentesusa.com/>

Real Academia Española. (2020). Software. *En Diccionario de la lengua española – Edición del Tricentenario*.

Recuperado de: <https://dle.rae.es/software>

Ruiz Salazar (1 de Junio de 2016). *¿Quieres Obtener Una Patente Para Tu Software?*

<https://www.ruizsalazar.cl/blog/quieres-obtener-una-patente-para-tu-software/> United States Patent and Trademark Office (s.f.) *USPTO Fee Schedule*. Department of Commerce. <https://www.uspto.gov/learning-and-resources/fees-and-payment/uspto-fee-schedule#Patent%20Fees>

Zárate, L. (2019). *El registro del software ante INDAUTOR en México y la posibilidad de patentamiento en algunos casos* (tesis de grado). INFOTEC Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación, Ciudad de México, México.



Esta obra está bajo una licencia de Creative Commons
Atribución-NoComercial-CompartirIgual 4.0 Internacional.