

TIES Revista de Tecnología e Innovación en Educación Superior

Publicación semestral • ISSN 2683-2968 • Noviembre 2023 • Número de revista 8
doi.org/10.22201/dgtic.26832968e.2023.8

Inteligencia Artificial para la solución de problemas sociales Alianza UNAM-Huawei

Editorial

Dra. María Josefa Santos Corral, Dra. Rebeca Nadia Ximena de Gortari Rabiela

Diseños globales prácticas locales en Inteligencia Artificial: el caso de la Alianza para el desarrollo de habilidades digitales

Diego Flores Jiménez

La creación de la Alianza, una perspectiva desde la UNAM

María Cristina Múzquiz Fragoso

GPU y NPU para el desarrollo de la inteligencia artificial. El caso NPU Ascend Huawei en Linux

José Fabián Romo Zamudio

Combatiendo la tuberculosis farmacorresistente desde la Inteligencia Artificial

Dora-Luz Flores Gutiérrez, Ricardo Perea-Jacobo, Guillermo Paredes Gutiérrez, Miguel Guerrero Chevannier, Raquel Muñoz Salazar

Plataforma de IA UNAM-HUAWEI para el mapeo de inundaciones mediante imágenes satelitales

Fernando Pech May, Jorge Magaña Govea, Luis Antonio López Gómez, German Ríos Toledo

Robot Autónomo para la detección y clasificación de fresas en campos de cultivo mediante Deep Learning

Luis Andrés Moreno Jiménez, Gesem Gudiño Mejía, Andrés Montes de Oca Rebolledo, Gerardo Flores

Inteligencia Artificial al servicio de los pequeños productores de dátil en México

Juan Pablo García Vázquez, Ricardo Salomón Torres, Dalila Blanca Pérez Pérez, Héctor Zatarain Aceves,
Jesús Eduardo Soto Vega, Edwin R. García Curiel



EDITORIAL

DRA. MARÍA JOSEFA SANTOS CORRAL
DRA. REBECA NADIA XIMENA DE GORTARI
RABIELA

Los avances en las investigaciones y aplicaciones de Inteligencia Artificial (IA) están generando grandes expectativas, no exentas de preocupación por las tareas humanas que están supliendo. A diario escuchamos noticias sobre cómo los nuevos chats; GTP, el Bing Chat y otros similares pueden hacer nuestros trabajos escolares, artículos académicos, controlar nuestras agendas, organizar viajes, componer música y hasta diseñar una nueva obra de arte.

Ante tal cambio que nos desborda y asombra no queda más que arremangarse y poner manos a la obra, para descubrir las ventajas y retos que ofrece esta nueva tecnología en nuestro quehacer cotidiano, tal como lo hicieron las sociedades que nos antecedieron cuando se concretaron otros grandes cambios tecnológicos.

La UNAM, el gobierno mexicano y las mismas empresas que ofrecen el software y hardware que posibilitan el diseño y uso de aplicaciones de IA, consideraron que cada uno de ellos tiene conocimientos que compartir y que pueden complementarse; con ello en mente es que negociaron un proyecto de transferencia de tecnología que se denominó "Alianza para Promover el Desarrollo de Capacidades Digitales en México". Éste es un espacio construido desde la lógica de la vinculación y la transferencia de conocimientos y tiene como objetivo el de integrar un paquete tecnológico que fomente el desarrollo de capacidades y aplicaciones de IA, para apoyar investigaciones que puedan, eventualmente, integrarse en desarrollos tecnológicos que resuelvan problemas sociales.

En este número de la revista TIES presentamos siete trabajos que muestran lo que ocurre en este espacio. Los primeros tres presentan las referencias bajo las cuales se integra la Alianza, los últimos cuatro dan cuenta de los resultados de los proyectos apoyados en la primera convocatoria emitida por la Alianza.

El trabajo Diseños globales prácticas locales en Inteligencia Artificial: el caso de La Alianza para el desarrollo de habilidades digitales, tiene como propósito definir los marcos institucionales globales y las preocupaciones locales, que permearon la negociación-construcción de la Alianza. El autor nos muestra también las referencias internacionales que guiaron las primeras negociaciones del Gobierno Federal para la cimentación de la Alianza y la manera en que se pensó la integración de los distintos actores que después la conformaron.

El segundo trabajo refiere la manera en que se integra el espacio de la Alianza, sus mecanismos de operación y cómo se acuerdan sus propósitos y metas. Así en el artículo La creación de la Alianza, una perspectiva desde la UNAM la autora explica la forma en que el Grupo Especial de Innovación (GEI), sentó las bases de la convocatoria y con ello los mecanismos de selección para delimitar que propuestas de investigación serían apoyadas con los recursos de la Alianza (el uso de los servidores, capacitaciones y becas).

En el último trabajo que integra este bloque La relevancia de las GPU y las NPU en el desarrollo e investigación de redes neuronales para la inteligencia artificial, el autor describe el papel que juegan las Unidades de Procesamiento Gráfico (GPUs) en los procesos de investigación y desarrollo para el diseño de nuevos modelos de redes neuronales. Estos presentan grandes ventajas para la investigación en IA entre las que se destacan: tiempos de entrenamiento más rápido, capacidad de procesamiento paralelo, procesamiento de datos a grandes escalas, etc.

En el segundo bloque, se presentan los resultados de cuatro de los proyectos apoyados en la primera convocatoria. Los trabajos se organizaron de manera en que se describiera el objetivo del proyecto, la forma en que se integró el equipo de trabajo, y los resultados que obtuvieron a partir de la participación del grupo de investigación en la convocatoria, entre los que se destaca, el contar con mayor capacidad y velocidad en el procesamiento de la información, integrar y/o visibilizar al grupo de trabajo, o hacer más conscientes los procesos de transferencia.

Así en el primer trabajo *Combatiendo la tuberculosis farmacorresistente: Un enfoque innovador utilizando inteligencia artificial en plataforma Huawei* los autores presentan los avances en el estudio de la farmacorresistencia y cómo y en qué la participación en el proyecto de la Alianza ha contribuido a los avances de sus investigaciones. En este sentido concluyen que la participación en el proyecto propicio el desarrollo de capacidades para procesar la información de datos genómicos, entre los integrantes del equipo; además de que el uso de espacio y poder de cómputo permitió generar información que apoye a la toma de decisiones clínicas.

El segundo trabajo *Plataforma de IA UNAM-HUAWEI para el mapeo de inundaciones mediante imágenes satelitales* refiere la manera en que el análisis de grandes datos derivados de la detección y mapeo de inundaciones y zonas de riesgo permite la creación de un modelo para la clasificación de zonas con presencia de inundaciones sobre las imágenes satelitales SAR de entrada. Como resultado de su participación en la Alianza el equipo integrado por profesores y alumnos reportó que los códigos fuentes de acceso libre proporcionado a través de la plataforma permitieron reducir el tiempo de procesamiento y por lo tanto el desarrollo del proyecto.

El trabajo *Robot Autónomo para la detección y clasificación de fresas en campos de cultivo mediante Deep Learning*, tercero en este bloque, tiene como propósito el diseño de un robot autónomo, que utiliza algoritmos de Deep Learning para detectar y clasificar fresas automáticamente según su grado de madurez. La participación en la convocatoria de la Alianza propició que el grupo se involucrara en una dinámica de cooperación empresarial-académica, logrando soluciones de inteligencia artificial aplicadas en la agricultura trajo.

El trabajo *Inteligencia Artificial al Servicio de los Pequeños Productores de Dátil en México*, último de este grupo, tiene como propósito desarrollar aplicaciones de IA que puedan apoyar las actividades de selección y clasificación de los pequeños productores de dátil. Gracias a su participación en el proyecto, el grupo de investigación logró establecer nuevos lazos de colaboración entre investigadores de universidades y de los sectores productivos del país, así como desarrollar capacidades digitales y formar recursos humanos en el área de IA. Lo anterior, señalan los propios investigadores, tendrá un fuerte impacto social a mediano y largo plazo, beneficiando al sector primario del México.



TIES Revista de Tecnología e Innovación en Educación Superior

DISEÑOS GLOBALES PRÁCTICAS LOCALES EN INTELIGENCIA ARTIFICIAL: EL CASO DE LA ALIANZA PARA EL DESARROLLO DE HABILIDADES DIGITALES

DOI: 10.22201/dgtic.26832968e.2023.8.1

Diego Flores Jiménez (diegofloresjimenez@gmail.com)
*Consultor en temas de transformación digital y relación
con gobierno.*
Ciudad de México, México.

www.ties.unam.mx

Fecha de recepción: julio de 2023 • Fecha de publicación: noviembre, 2023

Noviembre 2023 | número de revista 8 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

DISEÑOS GLOBALES PRÁCTICAS LOCALES EN INTELIGENCIA ARTIFICIAL: EL CASO DE LA ALIANZA PARA EL DESARROLLO DE HABILIDADES DIGITALES

Resumen

El avance tecnológico acelerado, el impacto de las nuevas tecnologías y las prioridades nacionales (soberanía digital) son factores que han impulsado la búsqueda y el establecimiento de mecanismos de cooperación multiactor encaminados al positivo desarrollo tecnológico. Este último aspecto, implica que en las conversaciones globales se considere como prioridad a las personas, es decir, ellas deben ser el centro. En este sentido, el bienestar social digital y la inclusión son bases para que la generación de nuevas tecnologías pueda aportar soluciones y mejoras a las comunidades, así como a su contexto. La alianza para el desarrollo de habilidades digitales entre la UNAM-SRE-Huawei es ejemplo de buena práctica, la cual debe y puede compartirse -y replicarse- a nivel nacional e internacional.

Palabras clave:

Nuevas tecnologías, bienestar social digital, inclusión digital, transformación digital, habilidades digitales, soberanía tecnológica, cooperación triple hélice.

GLOBAL DESIGNS LOCAL PRACTICES IN ARTIFICIAL INTELLIGENCE: THE CASE OF THE ALLIANCE FOR THE DEVELOPMENT OF DIGITAL SKILLS

Abstract

The rapid technological change, the impact of new technologies and national priorities (digital sovereignty) are factors that have driven the search and establishment of multi-stakeholder cooperation mechanisms aimed to develop a positive technological development. This implies that in global conversations people are considered a priority, that is, they must be the center. In this matter, digital social well-being and inclusion are the foundations for the generation of new technologies to provide solutions and improvements to communities, as well as their context. The alliance for the development of digital skills between UNAM-SRE-Huawei is an example of good practice, which should and can be shared -and replicated- nationally and internationally.

Keywords:

New technologies, digital social well-being, digital inclusion, digital transformation, digital skills, technological sovereignty, triple helix cooperation.

DISEÑOS GLOBALES PRÁCTICAS LOCALES EN INTELIGENCIA ARTIFICIAL: EL CASO DE LA ALIANZA PARA EL DESARROLLO DE HABILIDADES DIGITALES

Introducción

La Alianza para el desarrollo de habilidades digitales se genera y desarrolla en el contexto de la transformación digital en México que ha tenido un enfoque en: i) soberanía tecnológica; ii) conectividad-infraestructura, al igual que; iii) el uso y desarrollo de software libre, así como de estándares abiertos.

Dado lo anterior, se ha propuesto que la transformación digital tenga una perspectiva humana. Es decir, que la inclusión digital sea la base angular para que todos, incluidos los grupos que se encuentran en una situación de vulnerabilidad, puedan beneficiarse de esta transición global que está impactando cada aspecto de la vida cotidiana. Asimismo, retoma la definición de inclusión digital: acceso equitativo, sustantivo, y seguro, al uso, al liderazgo y al diseño de las tecnologías digitales, servicios y oportunidades asociadas para todos, en todas partes [1].

Los consensos ya alcanzados en este sentido, en el marco de las Naciones Unidas, así como de la Unión Internacional de Telecomunicaciones (ITU, por sus siglas en inglés) deberán servir como base para la construcción de una conectividad digital universal y sustantiva. Para lo anterior, es fundamental proponer una perspectiva transversal de nuevas tecnologías o de las tecnologías emergentes a partir del pensamiento exponencial y de innovación.

La transformación digital no solo puede pensarse en términos de eficiencia, sino en colocar al centro al

ser humano y su contexto. Así, la transformación digital inclusiva, se debe desarrollar desde lo local, y a la par, sumar los intereses y necesidades particulares.

Es indispensable pensar desde una óptica de bienestar social digital. En este orden de ideas, México busca promover la inclusión, es decir, reducir las desigualdades sociales y digitales, empoderar a los más vulnerables mediante la adquisición de habilidades digitales, y fomentar más iniciativas que permitan acercar y vincular al servicio público con la sociedad, así como mejorar su profesionalización.

La transformación digital debe de apuntar a:

- Digitalización para la recuperación post-covid. Identificar cuáles son los grupos desatendidos, así como las herramientas digitales necesarias para impulsar su recuperación económica y social postpandemia.
- Brecha Digital e Igualdad de Género. Identificar acciones de política pública enfocadas en cerrar las brechas de género digitales y asegurar la participación de las mujeres en el ecosistema digital.
- Accesibilidad digital para personas en condición de vulnerabilidad. Diseñar y desplegar dispositivos y plataformas que disminuyan las brechas en el internet, con base en sus necesidades, incluidas las personas en movimiento (migrantes o desplazados); que se autoidentifican como indígenas o de pueblos originarios, entre otras.

Desde México, observamos que el proceso de transformación digital, ineludiblemente, debe contemplar la soberanía tecnológica de los países. Así pues, debe entenderse como una atribución única del Estado en la toma de decisiones, sin interferencias externas sobre qué política o estrategia se debe seguir en el ámbito digital y tecnológico.

En cuanto a los enfoques que desde el gobierno se deben tomar frente a las tecnologías emergentes resaltan las siguientes:

la necesidad de adaptar las nuevas tecnologías —mediante marcos institucionales y regulatorios— para que nuestro progreso sea sostenible;

- elevar los estándares educativos y las capacidades tecnológicas en nuestra juventud;
- extender la infraestructura digital de manera asequible para todos;
- atender las implicaciones éticas, culturales y socioeconómicas, así como la pérdida de empleos y desigualdades que conlleva el cambio tecnológico;
- asegurar el equilibrio entre la protección de la privacidad y el intercambio de datos;
- forjar una economía del conocimiento basada en emprendimientos de alto impacto, así como ecosistemas de innovación inclusivos, mediante una cooperación de quintuple hélice¹, a la par de métricas de inclusividad digital para la medición de impacto.

Surgimiento de la Alianza

Es en este contexto, que surgió la Alianza la cual tiene su origen en la conceptualización del fenómeno exponencial del cambio tecnológico, y como se logran establecer mecanismos de cooperación. Es decir, busca innovar en la forma en que pueden detonarse vinculaciones exitosas, al comprender que como punto de partida se requieren de los distintos sectores de la sociedad para diseñar soluciones a los retos que plantean las nuevas tecnologías con su desarrollo acelerado.

En ello, es fundamental que las tecnologías pongan al centro de su desarrollo a la persona, y no a la eficiencia, permitiendo generar; i) inclusión; ii) reducción de brechas sociales tradicionales y digitales, a la par de; iii) el establecimiento del bienestar social digital, que

¹ Por quintuple hélice se refiere a modelos de innovación, que tienen como punto de partida las esferas de colaboración en torno a Economía, Gobierno, Academia, Organizaciones de la Sociedad civil y medio ambiente. Esto porque el factor de sostenibilidad es toral en el avance tecnológico y social.

engloba temas como inclusión financiera, incorporación a la economía digital, desarrollo de habilidades digitales, higiene digital, privacidad y datos personales, y el internet como derecho humano para el desarrollo.

Respecto de la inclusión digital, es importante destacar que con base en el informe del enviado especial para Tecnología del Secretario General de las Naciones Unidas titulado “Achieving universal and meaningful digital connectivity. Setting a baseline and targets for 2030”, debe considerar al menos seis elementos habilitadores de conectividad: i) infraestructura; ii) asequibilidad; iii) dispositivos; iv) habilidades, así como, v) medidas de seguridad y protección [2].

Asimismo, siguiendo las conversaciones globales en foros multilaterales, principalmente en las Naciones Unidas el Secretario General, Antonio Guterres, ha hecho de esto una prioridad, al incrementar los diálogos y las acciones concertadas para comprender de mejor manera el cambio tecnológico exponencial, enfatizando los posibles impactos sociales, económicos, culturales y en la vida diaria de las personas en el mundo. Derivado de ello, es que surgió en 2016, el Panel en Cooperación Digital [3] que tiene como ente de acción la Hoja de Ruta en Cooperación de Digital [4] la cual incorpora diversas recomendaciones a las cuales el Enviado especial para Tecnología del secretario general da seguimiento.

En el ánimo de continuar su liderazgo en el tema de tecnología, derivado de la presentación de tres resoluciones, lideradas por México, ante la Asamblea General de la ONU en 2017, 2018 y 2021, es que nuestro país solicitó encabezar la recomendación 1.C sobre inclusión digital en conjunto con ONU Mujeres y Canadá, de la cual se obtuvo como resultado, el establecimiento de las bases y objetivos para alcanzar la conectividad digital para 2030 [2].

Bajo esta óptica, que comprende una visión interinstitucional e intersectorial, es que la Secretaría de Relaciones Exteriores (SRE), y la entonces Jefatura de Oficina de la Presidencia de la República, mediante el Consejo Nacional para el Fomento a la Inversión, al Empleo y el Crecimiento Económico (COFINECE) se dieron a la tarea de desarrollar proyectos de vinculación en materia de nuevas tecnologías, con el objetivo de que éstas tuvieran un impacto social y que la discusión global se reflejará en resultados que llegaran directamente a las personas, es decir, que los potenciales beneficios alcanzaran a las poblaciones y las comunidades.

La perspectiva de participación multipartita y multi-sectorial fue vital para la conformación de la Alianza, ya que es necesario contar con los recursos humanos para esta nueva revolución industrial basada en datos y procesamiento de información. En consecuencia, y atendiendo el mandato de la actual administración, es que se decidió buscar una institución de educación superior pública para generar en México un semillero de científicos de datos.

Es aquí cuando se decidió integrar como socio estratégico a la Universidad Nacional Autónoma de México (UNAM). El Rector Enrique Graue aceptó sumarse a la iniciativa, y nombró encargado del proyecto por parte de la Casa de Estudios al Dr. Héctor Benítez, director de DGTIC.

Asimismo, en las conversaciones sobre quiénes del sector privado podrían ayudar a hacer realidad esta Alianza, en materia de infraestructura y conocimiento práctico, y a consideración de COFINECE, se dio el acercamiento con Huawei, quién al entender y apropiarse de esta visión decidió sumarse a los esfuerzos, en el marco de su iniciativa TECH4ALL. Ésta última tiene su base en los Objetivos de Desarrollo Sostenible (ODS) de la ONU y tiende a priorizar el desarrollo, la inclusión, el género y hacer frente al cambio climático.

La Cooperación Alemana GIZ permitió incorporar el innovador proceso de design thinking² a los proyectos. Igualmente, el apoyo de la Secretaría del Trabajo y Previsión Social (STPS) mediante el programa “Jóvenes Construyendo el Futuro” permitió ampliar el impacto de las convocatorias y proyectos, al poder sumar más sectores y, sobre todo, una integración multidisciplinaria de los equipos de desarrollo e investigación que se adscribirían a las convocatorias de la Alianza.

Una vez que se contó con los principales integrantes fundadores de la Alianza es que comenzó, durante la pandemia de Covid-19, el diseño, desarrollo y articulación de un proyecto multi actor en nuevas tecnologías que fomenta la innovación, al buscar que la transformación digital genere conexiones más integradas e inteligentes —entre todas las cosas, dispositivos y personas— para no dejar a nadie atrás en el desarrollo de un bienestar social y digital.

Con esto en mente, y al comprender todas las necesidades nacionales, es que se define la misión de la Alianza, la cual cree firmemente que la construcción de una base de talento amplia e inclusiva con énfasis en

el desarrollo de aplicaciones para apoyar a los sectores vulnerables extiende el bienestar al resto de la población mexicana. Lo anterior, al hacer que los proyectos de investigación y desarrollo, y en general el de las tecnologías, sirvan al interés y al bien público, complementadas con una nueva ética social.

De ello derivan los principales objetivos de la Alianza en torno al bienestar social digital:

1. Promover el desarrollo y la instrumentación de las nuevas tecnologías para sembrar la inclusión y reducir las desigualdades sociales, al apoyar en especial a los sectores más vulnerables —tales como jóvenes desempleados, migrantes mexicanos retornados, micro emprendedores, indígenas y discapacitados, en particular a las mujeres,— por medio de la interacción con las instancias públicas a cargo de atenderlos.
2. Empoderar a los sectores más vulnerables mediante la adquisición de las habilidades digitales para un mayor bienestar.
3. Impulsar el desarrollo de los sectores vulnerables desde una alianza triple hélice —gobierno, sector educativo y científico, así como el privado— bajo la premisa integral del bienestar social digital.
4. Fomentar iniciativas novedosas para que las nuevas tecnologías detonen alternativas de acercamiento y vinculación del servicio público a la sociedad, buscando su mejora y profesionalización.
5. Contribuir a la integración de tecnologías de Inteligencia artificial (IA) y ciencia de datos en las microempresas con actividades de producción o servicios directos para los sectores más vulnerables.
6. Fomentar la capacitación avanzada en Tecnologías de la Información y las Comunicaciones (TIC) e IA.

La Alianza es un ejemplo de una mejor practica internacional y de cómo una propuesta surgida en foros internacionales tiene resultados reales y tangibles en poblaciones específicas. La culminación fue la inauguración del Laboratorio de Inteligencia Artificial en 2022 en Ciudad Universitaria, realizada por el rector, el secretario de Relaciones Exteriores y directivos de Huawei México.

¡Enhorabuena para fortalecer y promover el desarrollo de científicos de datos y tecnólogos mexicanos que nos permitan aprovechar las oportunidades que plantean las nuevas tecnologías! Que esta buena práctica, sirva para que, tanto en el presente, como en el futuro, se acentúe el acercamiento entre los hacedores de política pública, industria y comunidad científica.

² Design thinking es una metodología de diseño de resolución de problemas que permite desarrollar soluciones centradas en las personas.

BIBLIOGRAFÍA

[1] United Nations, “Digital Inclusion,” *United Nations Digital Inclusion*, 2021. [En línea]. Disponible en: https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/general/Definition_Digital-Inclusion.pdf

[2] International Telecommunication Union, “Achieving universal and meaningful digital connectivity: Setting a baseline and targets for 2030,” *United Nations International Telecommunication Union*, 2022. [En línea]. Disponible en: https://www.itu.int/itu-d/meetings/statistics/wp-content/uploads/sites/8/2022/04/UniversalMeaningfulDigitalConnectivityTargets2030_BackgroundPaper.pdf

[3] United Nations Secretary-General’s High-level Panel on Digital Cooperation, “The UN Secretary-General’s Roadmap on Digital Cooperation,” 2020. [En línea]. Disponible en: <https://www.un.org/en/sg-digital-cooperation-panel>

[4] United Nations, “Road map for digital cooperation: implementation of the recommendations of the High-level Panel on Digital Cooperation Report of the Secretary-General,” 2020. [En línea]. Disponible en: <https://www.un.org/en/content/digital-cooperation-roadmap/>

Fecha de recepción: julio, 2023

Fecha de publicación: noviembre de 2023

Cómo se cita:

D. Flores, “Diseños globales prácticas locales en Inteligencia Artificial: el caso de La Alianza para el desarrollo de habilidades digitales” *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 8, noviembre, 2023. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

LA CREACIÓN DE LA ALIANZA, UNA PERSPECTIVA DESDE LA UNAM

DOI: 10.22201/dgtic.26832968e.2023.8.2

María Cristina Múzquiz Fragoso (cristinamf@unam.mx)
*Universidad Nacional Autónoma de México, Dirección General
de Cómputo y de Tecnologías de Información y Comunicación.
Ciudad de México, México.*

www.ties.unam.mx

Fecha de recepción: julio de 2023 • Fecha de publicación: noviembre, 2023

Noviembre 2023 | número de revista 8 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

LA CREACIÓN DE LA ALIANZA, UNA PERSPECTIVA DESDE LA UNAM

Resumen

En 2020 nace una iniciativa entre Huawei, la Secretaría de Relaciones Exteriores y la UNAM, entre otras entidades, para apoyar a equipos de investigación en el campo de la inteligencia artificial (IA). Es así que se crea el Grupo Especial de Innovación. Por la UNAM participan diferentes entidades que dan un apoyo en diferentes sentidos a los investigadores. Al 15 de noviembre de 2023, se han lanzado dos convocatorias; una tercera está por ser emitida.

En cifras, se han premiado 26 proyectos y se ha capacitado a varias decenas de personas en temas de tecnología y de *Design Thinking*, así como en Inteligencia Artificial. Además, los investigadores han tenido avances en sus proyectos, gracias al equipo al que pueden acceder y que fue donado por la empresa Huawei.

Los proyectos ganadores tienen un objetivo común: apoyar a diferentes comunidades del país con un enfoque social. Las lecciones aprendidas han sido muchas y valiosas, en una iniciativa que se ha orientado a: “Consolidar una propuesta orientada a fortalecer el desarrollo, la investigación y la implementación de nuevas tecnologías, con énfasis en la inteligencia artificial (IA) y la inclusión en la economía digital, así como a detonar propuestas específicas para México”.¹

Palabras clave:

Alianza, Inteligencia Artificial, Capacitación, Proyectos.

THE ESTABLISHMENT OF THE ALLIANCE, UNAM’S PERSPECTIVE

Abstract

In 2020, Huawei, the Ministry of Foreign Affairs, and UNAM, among other entities, launched an initiative to support research teams in the field of artificial intelligence (AI). This initiative led to the establishment of the Special Innovation Group. UNAM entities actively participate in the initiative by providing support to researchers in various ways. Two calls for proposals have already been launched, and the third one is currently in development.

To date, 26 projects have been awarded, and several dozen individuals have received training in technology, design thinking, and artificial intelligence. The researchers have made significant progress in their respective fields, thanks in large part to the equipment donated by Huawei. The winning projects share a common goal of supporting diverse communities in the country with a social focus. The project has generated numerous valuable lessons learned.

The overarching objective of the initiative is to consolidate a proposal aimed at strengthening the development, research, and implementation of new technologies, with a specific emphasis on artificial intelligence (AI) and inclusion in the digital economy. The project also aims to generate specific proposals tailored to Mexico’s specific needs.

Keywords:

Alliance, Artificial Intelligence, Training, Projects.

¹“Alianza para promover el desarrollo de capacidades digitales en México”. <https://alianza.unam.mx/>

LA CREACIÓN DE LA ALIANZA, UNA PERSPECTIVA DESDE LA UNAM

Introducción

En 2020, Huawei y la Secretaría de Relaciones Exteriores (SRE) se plantearon impulsar una iniciativa para hacer uso de equipo de cómputo especializado propiedad de esa compañía y diseñado para desarrollar proyectos de Inteligencia Artificial. Para conformar un equipo de triple hélice, decidieron acercarse a la Universidad Nacional Autónoma de México (UNAM).

La discusión entre los equipos de la SRE, Huawei y la UNAM se centró en acordar un mejor uso de los recursos tecnológicos para impulsar proyectos en Inteligencia Artificial. Para lo cual, se propuso conformar un grupo académico denominado Grupo Especial de Innovación (GEI). Este se encargaría de publicar una convocatoria para dar a conocer la existencia de estos recursos y abrir la participación a grupos de investigación a través de la selección de los proyectos que podrían hacer uso de los nuevos equipos.

Con la conformación del GEI, se buscó:

- Crear un equipo multidisciplinario con la participación de expertos en Inteligencia Artificial y de otros especialistas con conocimiento sobre las necesidades sociales de las poblaciones más vulnerables del país.
- Generar un análisis objetivo para abordar con claridad las prioridades por atender en los proyectos de inteligencia artificial.
- Diseñar una evaluación rigurosa a partir del establecimiento de criterios y metodologías para analizar las

propuestas de proyectos, asegurando una evaluación justa y transparente.

- Promover una gestión eficiente de los recursos, en función de las necesidades de los proyectos.
- Evaluar los riesgos del proyecto, entre ellos, los tecnológicos, de usabilidad y laborales; además de hacer propuestas para atenderlos, considerando también los aspectos éticos y normativos.
- Ser transparente en la asignación de los recursos, conforme a lo establecido en las convocatorias y sus lineamientos.

Si bien la conformación del GEI ha cambiado a lo largo del tiempo, inicialmente estuvo integrado por:

- Universidad Nacional Autónoma de México, a través del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) y la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC)
- Huawei Technologies de México
- Secretaría de Relaciones Exteriores (SRE)
- Secretaría de Economía (SE)
- Consejo Nacional para el Fomento a la Inversión, el Empleo y el Crecimiento Económico (COFINECE)
- Agencia Alemana de Cooperación para el Desarrollo Sustentable (GIZ) en México
- Red de Soluciones para el Desarrollo Sostenible (SDSN) de la Organización de las Naciones Unidas (ONU).

Para potenciar al GEI y tener un equipo fortalecido, por la UNAM, se invitó al proyecto tanto al Instituto de Investigaciones Sociales (IIS) como a la Coordinación de Vinculación y Transferencia Tecnológica (CVTT). La incorporación del IIS tuvo como propósito generar una visión integral de los proyectos, sobre todo la comprensión sociotécnica de éstos. Mediante la integración de la CVTT, el objetivo fue apoyar la transferencia de conocimiento, tecnología y productos desarrollados a organismos y empresas del sector público, para intensificar su aprovechamiento y apoyar a los líderes de proyecto.

Así se definió un plan de trabajo a finales del 2020, que implicaba tener reuniones semanales para construir y publicar una convocatoria con los lineamientos para el uso de los equipos donados por Huawei.

Primera convocatoria

En el lanzamiento de la primera convocatoria y sus lineamientos, se tuvieron las siguientes consideraciones:

- Incluir a investigadores de todo México.
- Contar con propuestas de diseño para la aplicación de tecnologías o el desarrollo de productos factibles de transferir.
- Impulsar proyectos con un enfoque social.
- Ofrecer recursos atractivos a los equipos ganadores, los cuales consistieron en:
 - Accesos a recursos físicos (hardware) en seis servidores de diferentes capacidades.
 - Acceso a recursos lógicos (software): ejemplos de código fuente, a través de la comunidad Huawei-Ascend, y acceso a un foro de desarrolladores.
 - Capacitación técnica a todos los miembros de los equipos seleccionados. Se impartieron cursos en línea de carácter introductorio en temas como inteligencia artificial, machine learning y deep

learning; así como para el uso de la plataforma MindSpore que facilita el desarrollo de aplicaciones relacionadas con la inteligencia artificial, haciendo un mejor uso de los recursos energéticos.

- Curso de Design thinking, una metodología para desarrollar la innovación centrada en las personas, curso patrocinado por GIZ.
- Materiales en formato digital y documentos de referencia.
- Pruebas para obtener una certificación en línea sobre inteligencia artificial.
- Posibilidad de obtener certificaciones gratuitas, como: Huawei Certified ICT Associate (HCIA), Huawei Certified ICT Professional (HCIP) y Huawei Certified ICT Expert (HCIE).

En la convocatoria se tenían dos líneas de trabajo:

1. Organizar un foro con especialistas de alto nivel.
2. Realizar su difusión a nivel nacional.

Para la organización y difusión del foro, se integró un amplio directorio que incluyó a investigadores y directivos de las principales universidades del país, así como a grupos de la sociedad en general, vinculados con el sector académico, entre otros, además de personalidades destacadas en los campos de inteligencia artificial y tecnologías de la información.

El foro se celebró el 20 de abril del 2021 y tuvo como título: *Aplicaciones de Inteligencia Artificial para Proyectos de Impacto Social en el Marco de una Recuperación Sostenible post COVID-19*. Por una parte, participaron expertos que expusieron proyectos exitosos en inteligencia artificial, mostrando los retos que se enfrentaban; también se logró reunir a un gran número de personas interesadas en temas de IA, lo cual sirvió para anunciar el lanzamiento de la convocatoria. Las personas participantes en el foro fueron:

Líderes académicos	
Dr. Luis Enrique Sucar	Investigador del Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE).
Dr. Adolfo Guzmán Arenas	Profesor e investigador del Centro de Investigación en Computación (CIC) del Instituto Politécnico Nacional (IPN).
Dr. Fernando Arámbula	Investigador del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la UNAM.
Dr. Humberto Sossa Azuela	Profesor e investigador del Centro de Investigación en Computación del IPN y jefe del Laboratorio de Robótica de esa misma institución.
Dra. María Josefa Santos Corral	Investigadora del Instituto de Investigaciones Sociales (IIS) de la UNAM.

Líderes académicos	
Dra. Cristina Verde Rodarte	Investigadora del Instituto de Ingeniería (II) de la UNAM.
Dra. Pilar Ángeles	Investigadora del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la UNAM.

Líderes de gobierno	
Mtro. Francisco Medina Gómez	Director General del Consejo Estatal de Ciencia y Tecnología de Jalisco (COECYTJAL).
Lic. Lucía Meza Guzmán	Presidenta de la Comisión de Comunicaciones y Transportes del Senado de la República.
Mtro. Javier Juárez Mojica	Comisionado del Instituto Federal de Telecomunicaciones (IFT).

Líderes de la industria	
Liu Jiude	Director General de Huawei Technologies de México.

La convocatoria y sus lineamientos fueron publicados el 30 de junio del 2021; después de varios meses 10 proyectos fueron seleccionados para este primer ejercicio. Entre los temas sobresalientes estuvieron: la detección del nivel de madurez de las fresas mediante técnicas de aprendizaje profundo; la efectividad de las redes neuronales profundas en la selección de dátiles; formas de analizar el COVID-19; detección automática de inundaciones con imágenes satelitales; monitoreo ambiental; identificación y predicción de tuberculosis, utilizando métodos de machine learning; así como un prototipo de sistema web de asistencia médica para la clasificación de estudios mamográficos, utilizando métodos de aprendizaje profundo.

En este primer ejercicio nos percatamos de que hay mucho trabajo por realizar en relación con la IA, aún cuando no se trata de una disciplina nueva. Asimismo, identificamos diversos motivos relacionados con la baja recepción de proyectos en la convocatoria:

- Escasez de expertos en IA. Aún es insuficiente la capacitación en este tema y se requieren especialistas, lo cual implica recursos y tiempo.
- Desvinculación entre los equipos técnicos dedicados a programar algoritmos de inteligencia artificial y los expertos en temas sociales. Los proyectos para este tipo de convocatorias requieren de equipos multidisciplinarios.
- Falta de infraestructura que soporte la formación de las personas. Usualmente se requieren equipos especializados a los cuales no se tiene acceso.

- Dificultad para contar con datos de calidad que permitan entrenar y evaluar los modelos. Algunos de los investigadores tenían muy buenas ideas, pero carecían de datos para trabajar.
- Temporalidad en el uso de los equipos. Algunos investigadores mencionaron que los proyectos de IA requieren años de trabajo y la convocatoria sólo los apoyaba por un periodo reducido.
- Desconfianza sobre los derechos de autor. Si bien la convocatoria dejaba claro que los derechos de la investigación eran del equipo, hubo incredulidad al respecto.

Acerca de los equipos, se debe señalar que entre los líderes de los proyectos solo había una mujer. Al respecto cabe decir que hay una escasez de mujeres en las carreras y profesiones STEM (ciencias, tecnología, ingeniería y matemáticas). En el mundo de las Tecnologías de Información y Comunicación, el porcentaje de su participación en estas carreras se ubica entre el 10 y 30%². La explicación de este fenómeno es multicausal, sin embargo, algunos estudios sugieren que entre mayores referentes exitosos femeninos tengan las niñas y adolescentes, mayor será su participación en las tecnologías.

² En los cuadernos de Planeación Universitaria - Año 2022, de la UNAM, se reporta el siguiente porcentaje de egresadas mujeres en las siguientes carreras: Ingeniería en computación (23.5%), Ingeniería en Telecomunicaciones (18%), Ciencias de la computación (25%), Matemáticas (31%), Física (20.2).

Uno de los retos más sobresalientes de la primera convocatoria fue el uso de los equipos donados. Debido a la falta de información sobre su manejo, a que responden a un ecosistema de trabajo totalmente distinto y al idioma de los manuales de operación, fue necesario que especialistas del IIMAS y la DGTIC realizaran una inmersión en esta tecnología para capacitar a los investigadores de los proyectos. Este proceso requirió mucho tiempo de aprendizaje y trabajo.

Segunda convocatoria

Con las lecciones aprendidas, el GEI publicó una segunda convocatoria. En esta se formalizó la capacitación impartida durante el primer ejercicio. Además se contó con becas del programa Jóvenes Construyendo el Futuro y becas otorgadas a través de la UNAM, patrocinadas por exalumnos.

Al igual que en la primera experiencia, se organizó un foro, ahora llamado Foro de Inteligencia Artificial. Ahí se anunció la nueva convocatoria y se firmó un convenio entre la UNAM, Huawei y la SRE; además se inauguró el Laboratorio de Inteligencia Artificial y Alta Tecnología en el IIMAS. Esto ocurrió el 2 de agosto de 2022, en Ciudad Universitaria.

A este acto fueron invitados tres líderes de la primera convocatoria, quienes mencionaron las diferencias entre los equipos convencionales y los del nuevo laboratorio, obtenidos a través de la Alianza. Una de las principales ventajas de los nuevos equipos estaba relacionada con los tiempos de procesamiento, que eran mucho menores, lo que les permitía obtener resultados más ágiles. De igual forma hicieron mención del curso *Design Thinking*, impartido por GIZ México, consideraron que la principal aportación del curso fue detectar que se enfocaron en el desarrollo de un nuevo producto, pero no se habían preocupado por las necesidades de los usuarios finales. Al finalizar el curso, pudieron contactar a usuarios potenciales y colaborar con ellos, lo que en muchos de los casos los llevó a un replanteamiento del proyecto, ya atendiendo problemas que los usuarios querían resolver.

A partir de los aprendizajes obtenidos en la primera convocatoria y con la retroalimentación de los investigadores, se publicó la segunda convocatoria; en esta ocasión se logró la participación de 16 proyectos.

La tercera convocatoria se publicará el 27 de noviembre del 2023 y estará abierta a más proyectos de investigación; no estará limitada a trabajos orientados a la solución de problemáticas sociales.

BIBLIOGRAFÍA

[1] "Alianza para promover el desarrollo de capacidades digitales en México". <https://alianza.unam.mx/>

[2] "Cuadernos de Planeación Universitaria - Año 2022. Perfil de Alumnos Egresados del Nivel Licenciatura". octubre de 2020. [En línea]. Disponible en: https://www.planeacion.unam.mx/ee/Publicaciones/pdf/perfiles/egresados/p_eg2020-2021.pdf

Fecha de recepción: julio, 2023

Fecha de publicación: noviembre de 2023

Cómo se cita:

M. C. Múzquiz, "La creación de la Alianza, una perspectiva desde la UNAM," *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 8, noviembre, 2023. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

**GPU Y NPU PARA EL DESARROLLO DE LA INTELIGENCIA
ARTIFICIAL. EL CASO NPU ASCEND HUAWEI EN LINUX**

DOI: 10.22201/dgtic.26832968e.2023.8.3

José Fabián Romo Zamudio (fabian.romo@unam.mx)
*Universidad Nacional Autónoma de México, Dirección
General de Cómputo y de Tecnologías de Información y
Comunicación. Ciudad de México, México.*
ORCID: 0009-0003-9269-8185

www.ties.unam.mx

Fecha de recepción: julio de 2023 • Fecha de publicación: noviembre, 2023

Noviembre 2023 | número de revista 8 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

GPU Y NPU PARA EL DESARROLLO DE LA INTELIGENCIA ARTIFICIAL. EL CASO NPU ASCEND HUAWEI EN LINUX

Resumen

Los rápidos avances en Inteligencia Artificial (IA) se han potenciados por el desarrollo de hardware de computadora más capaz y algoritmos innovadores. Las Unidades de Procesamiento Gráfico (GPU) emergieron como un componente esencial para la investigación y el desarrollo de la IA, otorgando un inmenso poder computacional para acelerar el entrenamiento y la inferencia de las redes neuronales. Este artículo se adentra en el uso de las GPU para la investigación y desarrollo de nuevos modelos de redes neuronales, partiendo de los esquemas más conocidos, en lo que corresponde a los proyectos del espacio de innovación para el desarrollo de habilidades digitales Alianza UNAM – Huawei, con un enfoque específico en los servidores Linux que contienen las Unidades de Procesamiento Neuronal (NPU) Ascend de la marca Huawei y sus capacidades de aceleración de los cálculos para la IA.

Palabras clave:

Arquitectura DaVinci, Contenedor, GPU, Huawei, Inteligencia Artificial, Linux, Mindspore, NPU, Pytorch, Redes Neuronales, Software libre, TensorFlow, Unidad de Procesamiento Gráfico, Unidad de Procesamiento Neuronal.

GPU AND NPU FOR ARTIFICIAL INTELLIGENCE DEVELOPMENT. THE CASE HUAWEI'S ASCEND NPU ON LINUX

Abstract

The fast advancements in artificial intelligence (AI) have been largely fueled by the development of powerful computing hardware and innovative algorithms. Graphic Processing Units (GPUs) have emerged as a vital component in AI research and development, providing immense computational power to accelerate neural network training and inference. This article delves into the usage of GPUs for AI research and the development of new neural network models, starting from the most well-known schemas regarding the projects under the space for innovation for digital capabilities development inside the UNAM-Huawei Alliance, with a specific focus on Linux-based servers and the Huawei NPU Ascend in supporting AI workloads and accelerating its computations.

Keywords:

DaVinci Architecture, Docker, GPU, Huawei, Artificial Intelligence, Linux, Mindspore, NPU, Pytorch, Neuronal Networks, Open source, TensorFlow, Graphic Processing Unit, Neuronal Processing Unit.

GPU Y NPU PARA EL DESARROLLO DE LA INTELIGENCIA ARTIFICIAL. EL CASO NPU ASCEND HUAWEI EN LINUX

1. Introducción

La Inteligencia Artificial (IA) ha revolucionado múltiples campos del saber, la producción y la educación al facilitar que los algoritmos para aprendizaje de máquina puedan abordar problemas complejos de muy diversos tipos. El crecimiento exponencial de la IA está íntimamente ligado a la disponibilidad de cómputo de alto rendimiento, capaz de manejar grandes volúmenes de datos y resolver operaciones con mayor demanda de cómputo. En este contexto, los microprocesadores conocidos como GPU¹ se han consolidado como elementos clave en el ámbito de la IA, proporcionando capacidades superiores, tanto computacionales como de procesamiento en paralelo. Adicionalmente, dos factores han incidido en ese desarrollo acelerado de la IA: la evolución de tecnologías para el aprendizaje profundo y la alta disponibilidad de grandes volúmenes de datos. Sin embargo, procesar esas cantidades de datos implica un entrenamiento de redes neuronales profundas con sustanciales recursos computacionales sustanciales, lo que de por sí es una labor matemáticamente intensa. Los microprocesadores tradicionales o CPU² son insuficientes para cumplir con las demandas del entrenamiento de esas redes neuronales complejas debido a su naturaleza secuencial y escalar. En consecuencia, las

GPU se han convertido en el mejor soporte de hardware para el entrenamiento de modelos de aprendizaje profundo, que son la base inobjetable de la inteligencia artificial actual. Las GPU, diseñadas inicialmente para la renderización, han sido por muchos años la mejor alternativa para acelerar los cálculos computacionales que requiere la IA.

Se procurará explorar aquí el papel fundamental que las GPU han tenido en la investigación en IA y el desarrollo de nuevos modelos de redes neuronales, enfocándose en el uso de las GPU en servidores con sistema operativo Linux y equipados con la tecnología de NPU Ascend desarrollada por la empresa Huawei, con la cual la Universidad Nacional Autónoma de México estableció un convenio de colaboración desde 2020 para crear el *Espacio de Innovación en el Desarrollo de Capacidades Digitales en México*³, a través del cual diversas instituciones mexicanas han podido ejecutar proyectos que incorporan tecnologías de IA para la atención de problemas en el país y el aporte de soluciones con un verdadero impacto social. Con base en ello, se procurará dar alguna orientación acerca del significado del uso de este tipo de tecnología de hardware en la mejora de proyectos y resultados de la inteligencia artificial, espe-

¹ GPU: Graphics Processing Unit. Unidad para Procesamiento de Gráficos.

² CPU: Central Processing Unit. Unidad de Procesamiento Central.

³ Para más información sobre los alcances y objetivos de la Alianza, consultar <https://alianza.unam.mx/>

cialmente los aprobados dentro de los proyectos de la Alianza establecida entre la UNAM y Huawei.

2. Las GPU en la investigación de IA

2.1. Arquitectura de las GPU y el cómputo paralelo

Las unidades de procesamiento gráfico son dispositivos altamente sofisticados para la ejecución de cómputo paralelo, y han experimentado un crecimiento significativo en su uso gracias a su capacidad para acelerar tareas computacionales complejas [1]. De manera general, una GPU se compone de los siguientes elementos que están diseñados para llevar a cabo cálculos paralelos de manera altamente eficiente:

- A) **Multiprocesadores de secuencias** (*Streaming Multiprocessors* o SM). representan los componentes fundamentales en la estructura de una GPU. Cada SM se compone de varias unidades de procesamiento conocidas como núcleos CUDA⁴ o núcleos de sombreado. La función primordial de los SM radica en la ejecución de instrucciones y el procesamiento de datos en un entorno paralelo.
- B) **Memoria global**. Representa el principal depósito de memoria en una GPU, donde se almacenan los datos accesibles por todas las SM. Aunque la Memoria Global tiene una capacidad considerable, su característica distintiva radica en una latencia relativamente elevada en comparación con otros tipos de memoria.
- C) **Memoria compartida**. Se presenta como un tipo de memoria más ágil y compacta localizada en cada SM. Su finalidad reside en facilitar una comunicación eficaz y en permitir el intercambio de datos entre los diversos subprocesos o hilos que operan en un mismo bloque.
- D) **Memoria de texturas y cachés**. Está optimizada para ubicar puntos en dos dimensiones y soportar el manejo de caché con el fin de tener ciclos de texturizado más eficientes. Adicionalmente, las GPU emplean diversos niveles de caché, tales como el L1 y el L2 para reducir la latencia en el acceso a la memoria.
- E) **Interfaz de memoria**. Facilita la comunicación entre la GPU y la memoria del sistema. Define el

⁴CUDA: Compute Unified Device Architecture. Es un tipo de plataforma de cómputo, propietaria de la empresa Nvidia, con la cual se pueden programar las GPU usando una variación del lenguaje C, llamado CUDA C.

ancho de banda y la latencia en las transferencias de datos con la memoria.

Por otra parte, el manejo eficiente de la memoria por las GPU es un elemento crítico para obtener resultados más precisos en tiempo considerablemente menor, a la vez que mejora los patrones de acceso a los datos y las transferencias de información. Además de las categorías de memoria global y compartida, las GPU operan con otras categorías que los investigadores en IA deben considerar al usar este tipo de procesadores [2]:

- A) **Memoria constante**. Es un espacio de memoria solo de lectura que almacena datos preservados de manera idéntica a lo largo de toda la ejecución de proceso. Con ello se puede tener acceso rápido y eficiente a valores constantes frecuentemente utilizados.
- B) **Memoria local**. Es el espacio de memoria reservado por subproceso o hilo, que se ubica dentro de la Memoria Global. Se emplea para almacenar variables locales y grupos de llamadas a funciones.
- C) **Archivo de registros**. Cada subproceso en un GPU tiene acceso a su propio archivo de registros, que se usan para almacenar variables y valores intermedios durante el cómputo. Es crucial optimizar el uso del registro para evitar fallas de anotación y mejorar el rendimiento.

2.2. Técnicas de aceleración por GPU para redes neuronales

Cuando se trata de acelerar el procesamiento y mejora de comportamiento de redes neuronales que usan las GPU, se pueden emplear varias técnicas:

- A) **Procesamiento por lotes**. Las GPU están especialmente diseñadas para procesar extensos conjuntos de datos de forma simultánea. Aumentar el tamaño de estos lotes conlleva el beneficio de mejorar el paralelismo y, por consiguiente, acelerar el tiempo de entrenamiento de manera general. No obstante, es esencial encontrar un equilibrio adecuado, ya que una elección de lote excesivamente grande podría ocasionar errores relacionados con los límites de memoria..
- B) **Gestión de la memoria de la GPU**. Las redes neuronales con modelos y conjuntos de datos de gran envergadura a menudo superan la capacidad de memoria de las GPU. Para mitigar este desafío, se requieren estrategias como el modelado recortado o la cuantificación por volúmenes diferenciados, que pueden significativamente reducir la carga en la me-

moria. Además, la adopción de técnicas de entrenamiento de precisión mixta, como el uso de FP16, puede disminuir aún más los requisitos de memoria adicionales, manteniendo un alto rendimiento.

- C) **Cómputo en paralelo.** Al utilizar técnicas de paralelización, tales como el paralelismo de datos y el paralelismo de modelos, se pueden distribuir las cargas de trabajo entre múltiples GPU o núcleos de GPU, reduciendo el tiempo de entrenamiento de la red neuronal que da solución al problema de Inteligencia artificial del que se trate. El *paralelismo de datos* implica replicar el modelo entre múltiples GPU y dividir los datos de entrada. El *paralelismo de modelos* involucra dividir en modelo entre varias GPU y con ello ejecutar los cálculos en paralelo.
- D) **Librerías y esquemas de programación para las GPU.** Al utilizar librerías y esquemas de programación optimizados se pueden obtener incrementos de velocidad en el entrenamiento de la red neuronal. Librerías como *cuDNN (CUDA Deep Neuronal Network)* facilitan implementar operaciones de aceleración que son comunes entre diversos tipos de redes neuronales, tales como las convoluciones. Los marcos de referencia para el aprendizaje profundo como *Tensorflow* y *Pytorch* se desarrollaron en ese sentido para mejorar la operación de las GPU.
- E) **Acumulación de gradientes.** En ocasiones, incluso después de aplicar técnicas de optimización, la memoria de la GPU podría seguir siendo insuficiente. Para abordar este desafío, se puede recurrir a la técnica de acumulación de gradientes. Aquí, los gradientes se calculan y acumulan tras la ejecución de pequeños lotes antes de llevar a cabo actualizaciones más significativas en los pesos del modelo. Esta estrategia posibilita el empleo posterior de lotes mucho más amplios y eficaces sin aumentar la demanda de memoria.
- F) **Núcleos de tensores.** Las modernas GPU incluyen la capacidad para trabajar con hardware especializado denominado “*núcleo de tensor*” que pueden acelerar las operaciones de matrices con precisión mixta. Estos núcleos de tensores tienen una mayor capacidad de salida de datos, combinada con una menor precisión matemática, lo cual deriva en entrenamientos e inferencias de los modelos más rápidos, sin sacrificar demasiado a la precisión.

- G) **Entrenamiento distribuido.** Cuando se trabaja con modelos o conjuntos de datos demasiado grandes, se puede utilizar el entrenamiento distribuido para usar múltiples GPU ubicados en varios nodos de cómputo. Tecnologías como *Horovod*, la Estrategia Distribuida de *TensorFlow* o la Capacidad de Paralelismo de Datos de *Pytorch* permiten el entrenamiento distribuido de forma eficiente, a la vez que se reduce el tiempo total de entrenamiento del modelo de red neuronal que sustente el proyecto de inteligencia artificial.

La eficacia de las técnicas mencionadas, previamente detalladas a los participantes de los proyectos aprobados en el marco de la Alianza UNAM – Huawei, está influida por diversos factores. Estos incluyen la arquitectura del modelo de red neuronal, el tamaño del conjunto de datos utilizado para nutrir el modelo y la configuración específica de los servidores dentro del Laboratorio del Espacio de Innovación. En las dos convocatorias publicadas hasta 2023, la mayoría de los líderes de proyecto optaron por configurar sus entornos de operación como contenedores (Docker). Esta elección les proporciona un mayor control sobre múltiples variables de la plataforma durante el entrenamiento de sus modelos de inteligencia artificial, sin depender exclusivamente de las configuraciones genéricas de los nodos de cómputo en el Laboratorio del Espacio de Innovación.

2.3. Ventajas de usar GPU en la investigación de IA

El uso de las unidades de procesamiento gráfico en la investigación en inteligencia artificial tiene diversas ventajas [3]. Entre las más importantes están:

- A) **Capacidad de procesamiento en paralelo.** Las GPU pueden acelerar significativamente el entrenamiento y la inferencia, comparados con las CPU tradicionales.
- B) **Tiempos de entrenamiento más rápidos.** Los modelos de aprendizaje profundo requieren, en ocasiones, un entrenamiento extenso con grandes volúmenes de datos. Las GPU permiten un entrenamiento más rápido. Este incremento en la velocidad permite a los investigadores iterar de forma más rápida sus modelos y experimentar con diferentes estructuras.
- C) **Complejidad extendida de los modelos.** Los investigadores pueden entrenar modelos con millones o miles de millones de parámetros, permitiéndoles resolver los retos que imponen las tareas de inteligencia artificial, con un rendimiento superior al promedio que se alcanza con las tradicionales CPU.

- D) **Técnicas para la optimización de los modelos.** Técnicas como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN) han demostrado un notable éxito en la manipulación de datos de imágenes y secuencias de datos, respectivamente. Gracias a la eficacia de las GPU, estas estrategias de optimización han propiciado una auténtica revolución en la investigación de la inteligencia artificial, influyendo de manera significativa en campos como la visión por computadora, el procesamiento del lenguaje natural y el reconocimiento de voz.
- E) **Procesamiento de datos a gran escala.** La investigación en inteligencia artificial a menudo implica el procesamiento de ingentes cantidades de datos, que pueden variar desde terabytes (TB) hasta petabytes (PB) en magnitud. Las GPU son capaces de afrontar con éxito las demandas computacionales asociadas con la tarea de procesar y analizar tales volúmenes de información. Esta capacidad se torna especialmente valiosa en aplicaciones que involucran redes neuronales de aprendizaje profundo, como el análisis de grandes conjuntos de imágenes o volúmenes extensos de texto.
- F) **Disponibilidad y eficiencia en costos.** Las GPU se encuentran ampliamente disponibles y son de fácil acceso para investigadores, presentando costos significativamente más bajos en comparación con hardware especializado, como las TPU (Unidades de Procesamiento de Tensores), en muchas de las tareas de investigación en inteligencia artificial. Además, debido al aumento en la demanda de GPU en diversas áreas, como los videojuegos y la minería de criptomonedas, la disponibilidad y los precios de las GPU se han vuelto más asequibles, convirtiéndolas en una elección cada vez más popular entre los investigadores en inteligencia artificial.
- A) **Flexibilidad y personalización.** El código abierto garantiza la flexibilidad y la capacidad de personalización, lo que habilita a los investigadores y desarrolladores en el campo de la inteligencia artificial a adaptar los entornos de cómputo según sus necesidades particulares.
- B) **Accesibilidad y disponibilidad.** Los populares esquemas de programación en inteligencia artificial, como *Tensorflow* y *PyTorch*, son de código abierto, lo que significa que cualquiera puede utilizar, modificar y contribuir a su evolución. Esta accesibilidad ha democratizado la investigación y el desarrollo en inteligencia artificial, al mismo tiempo que ha fomentado la colaboración y el intercambio de conocimientos entre la comunidad académica y la industria.
- C) **Comunidad de colaboración e innovación.** Numerosos proyectos de código abierto en el campo de la inteligencia artificial han progresado gracias a la colaboración intrínseca en la comunidad que los engendra. En este entorno, desarrolladores e investigadores se han unido, contribuyendo con código, documentación y mejoras a los entornos de programación. Este esfuerzo colectivo del software libre ha propulsado modelos altamente innovadores, algoritmos, redes neuronales y técnicas más allá de las limitaciones convencionales de la inteligencia artificial.
- D) **Transparencia y confianza.** El software libre fomenta la transparencia y la confiabilidad en los sistemas de inteligencia artificial. Al disponer del acceso al código fuente, investigadores y desarrolladores tienen la capacidad de examinar, verificar y validar los algoritmos y modelos empleados en las aplicaciones. Esto les permite identificar y abordar cualquier desviación, vulnerabilidad o preocupación ética, además de realizar auditorías independientes y revisiones por parte de expertos, asegurando así que los sistemas de inteligencia artificial sean responsables y confiables.
- E) **Adopción y estandarización.** Los ampliamente utilizados entornos de programación en el ámbito de la inteligencia artificial, como *Tensorflow* y *PyTorch*, han adquirido un estatus de facto como estándares en la industria. Estos proporcionan una plataforma unificada para el desarrollo de la inteligencia artificial, lo que a su vez fomenta la interoperabilidad y la colaboración en este campo.

3. Servidores basados en Linux para procesamiento de IA

3.1. El software libre en la IA

El software de fuente libre (*open-source*) ha tenido un impacto significativo en el desarrollo de la inteligencia artificial, al establecer un fundamento sólido que propicia la colaboración y la innovación. Algunos de los elementos más relevantes que sustentan el impacto de del software libre en la IA son los siguientes:

- F) **Integración con aceleradores de hardware.** Los esquemas de desarrollo de código abierto brindan respaldo y optimización para aceleradores de hardware, como las GPU y los NPU, permitiendo a investigadores y desarrolladores aprovechar de manera eficaz las capacidades de paralelismo de este tipo de hardware.

3.1. Linux en el desarrollo de la IA

Linux ha tenido un impacto crucial en la evolución de la inteligencia artificial, convirtiéndose en una elección destacada entre investigadores y desarrolladores en este ámbito. Entre las ventajas principales de Linux en el campo de la IA, destacan las siguientes:

- A) **Fuente abierta.** El código fuente de Linux se encuentra libremente disponible para los usuarios, lo que les permite examinarlo, modificarlo y distribuirlo a su conveniencia. Esta naturaleza de código abierto de Linux promueve la colaboración y la transparencia, lo que, a su vez, concede a los investigadores en inteligencia artificial la capacidad de personalizar y ajustar los entornos de cómputo a sus necesidades específicas. Esto incluye la posibilidad de modificar el núcleo del sistema operativo Linux (*kernel*) para optimizar las configuraciones del sistema y la integración más precisa de esquemas de programación y herramientas para la inteligencia artificial.
- B) **Flexibilidad y personalización.** Linux ofrece soporte para una amplia gama de arquitecturas de hardware, lo que resulta especialmente beneficioso para los desarrolladores, ya que les permite seleccionar los componentes que mejor se adapten a sus sistemas, optimizando así el manejo de las considerables cargas de trabajo relacionadas con la inteligencia artificial. El diseño modular de Linux posibilita a los usuarios la elección e instalación únicamente de los componentes necesarios, lo que se traduce en un desarrollo más eficiente y ágil.
- C) **Plataforma robusta y estable.** Linux es renombrado por su estabilidad, fiabilidad y robustez. Gracias a la excelencia del núcleo de Linux en la gestión de la memoria, la programación de procesos y las operaciones de entrada y salida, se logra una ejecución notablemente eficiente y precisa de todas las cargas de trabajo relacionadas con la inteligencia artificial.
- D) **Ecosistema más extenso.** Existe un amplio espectro de herramientas de código abierto, como bibliotecas,

marcos de programación y entornos de desarrollo, que desempeñan un papel fundamental en la evolución e investigación en el campo de la inteligencia artificial. Los entornos de inteligencia artificial más populares, como *Tensorflow* y *PyTorch*, como se mencionó previamente, están optimizados para su funcionamiento en sistemas Linux. Esto capacita a los desarrolladores para crear algoritmos de aprendizaje automático de última generación, diseñar arquitecturas de redes neuronales innovadoras y aprovechar modelos previamente entrenados. Además, las distribuciones de Linux facilitan la instalación y gestión de todo el software relacionado con aceleradores y otros componentes clave para la innovación en modelos de inteligencia artificial, gracias a sus manejadores de paquetes y repositorios.

- E) **Rendimiento y escalabilidad.** Linux se distingue por su excepcional rendimiento y escalabilidad, permitiendo aprovechar al máximo las capacidades del hardware moderno. Esto abarca desde CPU de múltiples núcleos y GPU hasta las NPU desarrolladas por Huawei, para garantizar el rendimiento óptimo en cargas de trabajo de inteligencia artificial. Asimismo, Linux ofrece una gestión eficiente de la memoria y la comunicación entre procesos, lo que capacita a los desarrolladores de inteligencia artificial para explotar plenamente los recursos computacionales disponibles y alcanzar los niveles de rendimiento más adecuados.

3.3. Distribuciones de Linux para el computo con GPUs

Están verificadas varias distribuciones de Linux para operar con GPU y NPU. En el caso particular del Laboratorio de Inteligencia Artificial UNAM – Huawei se usa Ubuntu versión 18.04, debido a que los controladores de las tarjetas con los procesadores NPU Ascend están validados en su funcionamiento con esa distribución y versión. Sin embargo, en el ámbito de Linux para el desarrollo de la IA, existen varias distribuciones con grados diversos de efectividad y compatibilidad con los circuitos aceleradores:

- A) **Ubuntu.** Está ampliamente recomendada para el desarrollo de la inteligencia artificial. Se puede tener una plataforma estable y confiable para la mayoría de las cargas de trabajo, compatible con los esquemas de programación de este ámbito, tales como *Tensorflow*,

Pytorch y *Scikit Learn*. Ubuntu dispone de ediciones especializadas como *Ubuntu Studio*⁵ y *Ubuntu Robotics*⁶, más enfocadas en satisfacer ciertas necesidades específicas para la inteligencia artificial.

- B) **CentOS**. Basada en el código fuente de Red Hat Enterprise Linux (RHEL), es una plataforma estable de nivel empresarial para el desarrollo de la inteligencia artificial. También tiene versiones con Soporte a Largo Plazo (LTS) que incluyen actualizaciones de seguridad, haciéndolo adecuado para el desarrollo de aplicaciones de inteligencia artificial en ambientes de producción. CentOS se enfoca fuertemente en la estabilidad y la confiabilidad, asegurando un rendimiento consistente, siendo compatible con la mayoría de los esquemas de desarrollo.
- C) **Fedora**. Distribución administrada por la comunidad y conocida por sus características de última generación con actualizaciones rápidas. En esencia tiene un balance adecuado entre la estabilidad y los paquetes de última generación, haciéndola recomendable también para el desarrollo de proyectos de inteligencia artificial. Fedora incluye un conjunto de herramientas de desarrollo y librerías para *Tensorflow*, *Pytorch* y *Caffe*. Más aún: contiene opciones para el uso de contenedores del tipo docker y kubernetes, muy recomendables para desarrolladores que requieren construir y ampliar las aplicaciones en este ámbito de operación. Varias investigaciones en el Espacio de Innovación tomaron como referencia para el funcionamiento de contenedores las plantillas creadas en Fedora de acceso libre.
- D) **Arch Linux**. Es una versión preferida por usuarios avanzados y desarrolladores. Tiene una base mínima para su instalación, lo que permite a los usuarios personalizar el ambiente de desarrollo para inteligencia artificial de acuerdo con sus necesidades específicas. También permite usar el repositorio para usuarios Arch (AUR) que contiene un rango amplio de paquetes de software para inteligencia artificial, mantenidos y desarrollados por la comunidad.
- E) **Deepin**. Esta es una distribución atractiva visualmente y muy amigable con el usuario, que se enfoca en proporcionar una experiencia bastante sencilla para quien lo emplea. Ofrece un entorno de escri-

torio adecuado para desarrolladores de inteligencia artificial que buscan emplear interfaces simples para el trabajo con las redes neuronales. Deepin incluye una cantidad importante de esquemas de desarrollo pre-instalados para inteligencia artificial, entre los que están *TensorFlow* y *Caffe*, haciéndolo conveniente para el desarrollo de inteligencia artificial prácticamente desde que se instala.

4. Unidades de Procesamiento Neuronal (NPU) Ascend de Huawei

4.1. Características generales de las NPUs Ascend

Las NPU Ascend de Huawei representan componentes de hardware altamente especializados diseñados para potenciar el procesamiento de cálculos en inteligencia artificial y mejorar el rendimiento de dispositivos Huawei, con especial enfoque en teléfonos inteligentes y otros dispositivos móviles [4]. Algunas de las características clave de estos procesadores incluyen:

- A) **Aceleración interconstruida para IA**. Los procesadores tradicionales no se concibieron para abordar de manera eficaz las exigencias computacionales de la inteligencia artificial. Por esta razón, Huawei ideó las NPU como aceleradores de hardware especializados, diseñados específicamente para ejecutar cálculos relacionados con la inteligencia artificial de manera altamente eficiente. Estas NPU se han integrado en la solución Kirin, que es un SoC⁷.
- B) **Arquitectura y diseño**. Aunque los detalles precisos acerca de los componentes que integran los procesadores de Huawei están protegidos por derechos de autor y son de naturaleza propietaria, se sabe que están especialmente orientados al procesamiento avanzado de datos. Esto se logra mediante la combinación de estructuras de memoria, algoritmos especializados adaptados al hardware y unidades de procesamiento que comparten similitudes con las GPU en su capacidad para manejar cargas de cálculo matemático. Los algoritmos desempeñan un papel esencial en la interacción con el hardware, particularmente en la resolución de redes neuronales de aprendizaje profundo. Por tanto, el diseño está centrado en mejorar el rendimiento, reducir el consumo de energía y optimizar la eficiencia en general.

⁵ <https://ubuntustudio.org/>

⁶ <https://wiki.ros.org/ROS/Tutorials>

⁷ SoC: System on a Chip. Corresponde a un circuito avanzado que incorpora en una sola tarjeta los elementos de una computadora.

- C) **Desarrollo y optimización.** El desarrollo de las NPU implica una sinergia entre el diseño de hardware, desarrollo de software y la optimización de componentes, una combinación considerada de manera integral por los ingenieros de Huawei. Un componente esencial de las NPU es su capacidad para gestionar operaciones matriciales, que resulta significativamente más eficiente en comparación con las GPUs tradicionales. Esto se debe a que las NPU están especializadas en resolver problemas relacionados con redes neuronales, lo que implica una disposición tridimensional de cálculos y, como resultado, una reducción en la cantidad de ciclos necesarios para obtener los resultados en comparación con otras arquitecturas de hardware.
- D) **Entrenamiento e inferencia.** Las NPU de Huawei son versátiles y admiten tanto el entrenamiento como la inferencia de modelos de inteligencia artificial. Durante la fase de entrenamiento, estas NPU aceleran las tareas que requieren un intenso poder de cómputo, lo que incluye la propagación hacia adelante y hacia atrás en el conjunto de datos, la actualización de los pesos específicos de la información y la optimización continua de los algoritmos que conforman las redes neuronales. En la etapa de inferencia, donde el modelo se especializa en predecir resultados basados en nuevos datos, las NPU de Huawei son capaces de ejecutar tareas de inteligencia artificial en tiempo real. Esto se aplica tanto en los servidores que componen el Laboratorio del Espacio de Innovación como en dispositivos periféricos.
- E) **Evolución y avances.** A lo largo de los últimos años, Huawei ha realizado numerosos avances en el desarrollo de las NPU. En cada iteración del hardware, se ha trabajado en la mejora del rendimiento, la reducción del consumo de energía y la ampliación del soporte de software, al mismo tiempo que se han optimizado los algoritmos de inteligencia artificial para enriquecer la investigación respaldada por las NPU.
- HiSilicon de Huawei, la arquitectura DaVinci de las NPU incluye varias características que la distinguen de las GPU y CPU ya descritas previamente en este artículo:
- A) **Alto rendimiento.** Las NPU DaVinci tienen amplias capacidades de procesamiento, lo que se estima mejor para la IA, ya que facilita el aprendizaje profundo y el entrenamiento de las redes neuronales a través del aprendizaje de máquina.
- B) **Versatilidad.** Esta arquitectura soporta un amplio rango de modelos de IA y marcos de desarrollo, lo que asegura la compatibilidad con los esquemas más populares (TensorFlow, Pytorch, Caffe, Keras). De esa manera los proyectos que se desarrollan en el espacio de innovación se pueden migrar de manera muy simple a otras plataformas de hardware y de software para extender las capas de entrenamiento e inferencia y con ello robustecer los modelos de IA que se están investigando.
- C) **Eficiencia energética.** Las NPU consumen lo mínimo de energía necesaria para maximizar el rendimiento por watt, lo que es adecuado no solo para el entrenamiento de grandes modelos de IA que procesan volúmenes monumentales de datos, sino también para que los modelos puedan ejecutarse en dispositivos más pequeños donde la capacidad y eficiencia energéticas son claves, tales como teléfonos inteligentes, cámaras y tabletas que componen parte del universo de la Internet de las Cosas.
- D) **Escalabilidad.** Esta arquitectura está diseñada para ampliarse en función del dispositivo y los escenarios de utilización. Ya sea en teléfonos, cámaras o centros de datos, la arquitectura se reduce o amplía conforme sea necesario, sin que se requiera una reprogramación de los algoritmos.
- E) **Aceleración de redes neuronales.** Se deja en la capacidad de las unidades especializadas de hardware y de los algoritmos avanzados la aceleración de las operaciones en la red neuronal, en ámbitos como las multiplicaciones matriciales o las convoluciones. Estas capacidades de aceleración incrementan la velocidad en las fases del entrenamiento y la inferencia de los modelos de IA que se aprobaron en el marco del Espacio de Innovación.
- F) **Flexibilidad en el procesamiento neuronal.** Las NPU de Huawei están diseñadas para proporcionar cálculos de punto fijo y de punto flotante, lo que permite procesar múltiples tipos de datos. Esta adapta-

4.2. Arquitectura y capacidades

La arquitectura empleada en las NPU que soportan los proyectos de la Alianza UNAM – Huawei se denomina DaVinci, y es una de las más avanzadas en su tipo para el desarrollo y aceleración de los cálculos que requiere la Inteligencia Artificial, desde los teléfonos inteligentes hasta los centros de datos. Desarrollada por el segmento

bilidad permite una utilización más eficiente de los recursos físicos mientras se atienden los requerimientos de las tareas de IA.

- G) **Inteligencia Artificial “en el dispositivo”**. Mucho del objetivo de las NPU reside en que se puedan ejecutar realmente algoritmos de IA en cualquier dispositivo, sin depender todo el tiempo de la conectividad a grandes centros de datos o los recursos en nube pública, a la vez que se protege la privacidad e integridad de los datos recolectados localmente.

5. Conclusiones

La creación del espacio de innovación dentro de la Alianza UNAM - Huawei ha permitido el uso de componentes especializados como lo son las unidades de procesamiento neuronal (NPU), las cuales son paso evolutivo desde las GPU. Los algoritmos de inteligencia artificial actuales son altamente demandantes de ciclos de procesamiento, especialmente del tipo matricial, para lo cual fueron diseñadas las NPU que se incorporan en los

nodos de cómputo del Laboratorio que sustenta el Espacio de Innovación. Los proyectos que fueron aprobados por las convocatorias de la Alianza han explotado estas capacidades de cálculo considerando elementos comunes a otros otras plataformas de investigación como lo son los esquemas de desarrollo (todos ellos apoyados en el lenguaje Python, el cual se ha convertido prácticamente en el estándar de facto para toda la innovación en inteligencia artificial) así también en redes neuronales que están disponibles a través de los repositorios con los que cuenta Huawei, desde donde se pueden descargar modelos de estas redes para adaptarlos a las necesidades específicas de los investigadores. Esta combinación de recursos de fuente abierta con tecnologías de hardware innovador ha impulsado el desarrollo de los proyectos y no solamente para la solución de problemas con enfoque social, como ha sido el objetivo de las convocatorias, sino también para la formación de recursos humanos especializados que han aprendido a utilizar de la mejor forma posible estas herramientas físicas y lógicas.

BIBLIOGRAFÍA

- [1] Nickolls, J., Buck, I., Garland, M., & Skadron, K. Scalable parallel programming with CUDA. *ACM Queue*, Vol. 6 issue 2. pp. 40-53. 2008. [En línea]. Disponible en <https://queue.acm.org/detail.cfm?id=1365500>
- [2]. Harris, M. J. Parallel prefix sum (scan) with CUDA. *NVIDIA Developer Technology*, Vol. 3 Issue 3, 1-8. 2007. [En línea]. Disponible en <https://developer.nvidia.com/gpugems/gpugems3/part-vi-gpu-computing/chapter-39-parallel-prefix-sum-scan-cuda>
- [3]. Owens, J. D., Luebke, D., Govindaraju, N., Harris, M., Krüger, J., Lefohn, A. E., & Purcell, T. J. A survey of general-purpose computation on graphics hardware. *Computer Graphics Forum*, Vol. 26 Issue 1, 80-113. 2007. [En línea]. Disponible en <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=2189271fd54b6980b2b20ef515ac114a02854219>
- [4]. Huawei Technologies Co., Ltd. *Huawei Atlas AI Computing Solution*. Online ISBN 978-981-19-2879-6. 2022. https://doi.org/10.1007/978-981-19-2879-6_6

Fecha de recepción: julio, 2023

Fecha de publicación: noviembre de 2023

Cómo se cita:

J. F. Romo, “GPU y NPU para el desarrollo de la inteligencia artificial. El caso NPU Ascend Huawei en Linux,” *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 8, noviembre, 2023. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

COMBATIENDO LA TUBERCULOSIS FARMACORRESISTENTE DESDE LA INTELIGENCIA ARTIFICIAL

DOI: 10.22201/dgtic.26832968e.2023.8.4

Dora-Luz Flores Gutiérrez (dflores@uabc.edu.mx)
ORCID: 0000-0002-2623-3772

Ricardo Perea-Jacobo (perear@uabc.edu.mx)
ORCID: 0000-0003-4164-7719

Guillermo Paredes Gutiérrez (paredesg@uabc.edu.mx)
ORCID: 0000-0002-8519-7126

Miguel Guerrero Chevannier
(miguel.angel.guerrero.chevannier@uabc.edu.mx)
ORCID:0000-0002-2545-1209

Raquel Muñoz Salazar (ramusal@uabc.edu.mx)
ORCID: 0000-0001-5528-7772

*Todos los autores están adscritos a la Universidad
Autónoma de Baja California. Mexicali, Baja
California, México.*

Fecha de recepción: Junio 2023 • Fecha de publicación: noviembre, 2023

www.ties.unam.mx

Noviembre 2023 | número de revista 8 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

COMBATIENDO LA TUBERCULOSIS FARMACORRESISTENTE DESDE LA INTELIGENCIA ARTIFICIAL

Resumen

La tuberculosis (TB) es una enfermedad infecciosa importante para la salud global. El incremento de casos farmacorresistentes ha impactado en el costo y el tiempo de tratamiento de esta enfermedad. Es urgente contar con métodos de diagnóstico y clasificación de la farmacorresistencia con tiempos de respuesta rápida y precisa, para que de esta manera se controle, se trate y se disminuya la transmisión de la TB en la comunidad. Con la creciente cantidad de cepas resistentes y el acceso a datos genéticos y clínicos de estos casos se presenta la oportunidad de innovar con tecnología de inteligencia artificial en la detección temprana de casos de TB farmacorresistente.

Palabras clave:

Drogorresistencia, tuberculosis, inteligencia artificial, Huawei, diagnóstico.

COMBATING DRUG-RESISTANT TUBERCULOSIS FROM ARTIFICIAL INTELLIGENCE

Abstract

Tuberculosis (TB) is a major global health issue caused by an infectious disease. The rise in drug-resistant cases has impacted the cost and duration of treatment for this disease. It is crucial to quickly and accurately diagnose and classify drug-resistant strains to control, treat, and prevent the transmission of TB in the community. As more resistant strains emerge and genetic and clinical data become available, there is potential to utilize artificial intelligence technology in detecting drug-resistant TB cases early on.

Keywords:

Drug resistance, tuberculosis, artificial intelligence, Huawei, diagnosis.

COMBATIENDO LA TUBERCULOSIS FARMACORRESISTENTE DESDE LA INTELIGENCIA ARTIFICIAL

Introducción

La tuberculosis (TB) es una enfermedad infecciosa causada por bacterias del complejo *Mycobacterium tuberculosis*. La TB afecta principalmente los pulmones, pero también puede presentarse entre otros sitios (sistema nervioso, huesos, piel, intestinos, genitales, ganglios, entre otros). Aproximadamente, un tercio de la población mundial está infectada por la bacteria de *M. tuberculosis*. En 2021, 10 millones de personas enfermaron de TB y 1.6 millones murieron por esta enfermedad, lo que equivale a 4,500 decesos por día. Por lo anterior, la TB es considerada una emergencia de salud global [1].

La Organización Mundial de la Salud (OMS), estimó que en el 2021 se presentaron a nivel global 558,000 casos (483,000 a 639,000) de TB resistente a rifampicina (TB-RR), el fármaco de primera línea más efectivo. De estos, el 82% era TB multi-farmacorresistente (TB-MDR), es decir, también presentaba resistencia a por lo menos al segundo fármaco más importante, isoniacida. La farmacorresistencia se ha incrementado en los últimos 15 años, llegando a generar cepas de *M. tuberculosis* que son resistentes a todos los fármacos existentes para tratar a la enfermedad, denominándose cepas extremadamente farmacorresistentes (TB-XDR) [1]. La TB farmacorresistente (TB-DR), se transmite de la misma forma que la TB sensible. Sin embargo, es complicado tratar y curar la TB-DR, debido a que los fármacos son más caros y tóxicos que los fármacos utilizados para la TB sensible.

La estrategia tradicional para el diagnóstico de la TB-DR se basa en las pruebas de sensibilidad farmacológica a partir del cultivo microbiológico. La principal limitación es que se requieren de 12 a 16 semanas para obtener resultados, tiempo en el cual el paciente, al no tener el tratamiento farmacológico correcto, seguirá transmitiendo la bacteria entre la población. A nivel mundial, solo el 64% de los casos de TB son diagnosticados, es decir, de los 10 millones de personas que se diagnostican por primera vez, 3.6 millones de personas no saben que están enfermas de TB [1]. Hoy en día, los métodos moleculares tienen ventajas considerables para escalar la gestión y vigilancia programática de la TB-DR, ofreciendo diagnósticos rápidos, eficientes y sin el requisito del uso de un área de bioseguridad nivel 3 [2]. La OMS desde el 2010 ha recomendado el uso de pruebas moleculares rápidas y sensibles con el fin de reemplazar o complementar las pruebas convencionales existentes para detectar el complejo *M. tuberculosis* y resistencia a rifampicina [3].

La secuenciación del ADN con métodos de nueva generación (SNG) detecta rápidamente mutaciones asociadas con la resistencia a fármacos contra la TB, lo que le permite al médico tener la información necesaria para implementar el tratamiento farmacológico de manera individualizada, oportuna y expedita [3]. Consecuentemente la cadena de transmisión de la enfermedad se cortaría de manera más temprana disminuyendo la ge-

neración de nuevos casos. Sin embargo, la información de la SNG se debe de procesar con modelos computacionales complicados, de los que solo los expertos en bioinformática pueden analizarlos, y el análisis puede tardar días. Por lo que, es imperativo desarrollar interfaces o implementaciones amigables para los usuarios y tomadores de decisiones. Escalando las implementaciones para mantener la velocidad de procesamiento, rendimiento y costo computacional al alcance de las instituciones de salud.

La inteligencia artificial (IA), se refiere a la capacidad del sistema para interpretar datos externos correctamente, aprender de dichos datos y usar esos aprendizajes para lograr objetivos y tareas específicas utilizando una adaptación flexible [4]. La IA está transformando la práctica médica, la utilidad de la IA se ha estudiado en múltiples áreas de la salud, incluyendo la medicina de precisión, salud poblacional, procesamiento de imágenes en radiología, dermatología y oftalmología, entre otros [5]. La IA permite analizar grandes conjuntos de datos digitales disponibles para generar modelos idóneos de capacitarse a sí mismos en una tarea específica [4]. La IA aplicada en modelos de cómputo permite determinar el perfil de farmacorresistencia de la bacteria con un costo menor en tiempo y con una precisión mayor a los métodos tradicionales fenotípicos. El manejo de la TB-DR requiere del uso de herramientas de diagnóstico con una alta sensibilidad y especificidad, y algo muy importante, que los resultados puedan obtenerse en cuestión de horas, permitiendo de esta manera, que el paciente pueda iniciar el tratamiento de manera inmediata, cortando la cadena de transmisión [3], [6].

El proyecto “Identificación y predicción de farmacorresistencia en pangenoma de *Mycobacterium tuberculosis* utilizando métodos de Machine Learning”, que estamos desarrollando en alianza con UNAM-HUAWEI, tiene el objetivo de predecir la farmacorresistencia a partir de la información del genoma completo de aislados clínicos de *M. tuberculosis* mediante un modelo de red neuronal convolucional reduciendo el tiempo de diagnóstico para disminuir la incidencia de pacientes que desarrollen farmacorresistencia.

Avances en el estudio de la farmacorresistencia

Muchos países, incluyendo México, dependen de la baciloscopía para diagnosticar TB, prueba que viene utilizándose desde hace más de 100 años, sin embargo, esta

técnica no detecta a la TB-DR. La farmacorresistencia en México no se realiza de manera rutinaria, sólo en aquellos casos clasificados como recaída y o fracasos al tratamiento. La técnica más común es la Prueba de Sensibilidad Farmacológica fenotípica utilizando cultivo líquido, y se prueban con los cuatro fármacos de primera línea (rifampicina, isoniazida, etambutol y pirazinamida). Recientemente, la técnica de Xpert MTB/RIF se incorporó como técnica de diagnóstico en el Programa Nacional de Tuberculosis, pero solo se aplica a muestras que cumplan con ciertos criterios [7]. Sin embargo, los estudios de SNG sólo se han realizado de manera aislada y principalmente con propósitos de investigación. Al momento solo existen 1,208 genomas completos de aislados de *M. tuberculosis* de pacientes diagnosticados con TB en México. De los cuales solo el 7.2% (87) son de *M. tuberculosis* y el resto de *M. bovis* (92.8%, 1,121). Sin embargo, los genomas de *M. tuberculosis* tienen mayor representación de México, ya que provienen de 17 de los 32 estados de México, mientras que *M. bovis* solo son de 10 estados [8], [9], [10], [11], [12], [13], [14].

Herramientas para abordar el estudio de la farmacorresistencia

Día a día, se tiene acceso a mayor cantidad de datos biomédicos, particularmente a información de secuencias de genomas bacterianos. Sin embargo, a pesar de que la OMS ha recomendado el uso de tecnologías de IA, la mayoría de los Sistemas de Salud a nivel mundial, no cuentan con acceso fácil a ellas, debido a su alto costo y mantenimiento [15], [16].

Actualmente, la exploración del genoma completo de *Mtb* se encuentra en la frontera de la investigación de la farmacorresistencia, esto debido a que las tecnologías actuales solamente detectan pocas mutaciones para un número limitado de fármacos. En 2021, la OMS publicó un catálogo de mutaciones que incluye más de 17,000 mutaciones, de las cuales en menos del 2% se ha logrado conseguir suficiente evidencia para poder catalogarlas como asociadas a TB-DR [17]. La limitante actual, es la falta de capacidad estadística para realizar estos análisis, tanto por el alcance de la potencia de cómputo para el manejo de datos como los espacios para realizar los análisis. Actualmente, las investigaciones que utilizan soluciones de IA analizan simultáneamente una gran cantidad de mutaciones en intervalos de 500,000

- 650,000, por lo que el acceso a servidores y soluciones de IA, brindan un apoyo importante a la investigación biomédica [18], [19], [20].

El acceso a los servidores Huawei es un punto de partida para la colaboración de instituciones públicas y privadas con orientación social, para facilitar y acelerar el desarrollo de herramientas y soluciones que requieren los sistemas de salud nacional. Las plataformas Huawei, con su gran capacidad de procesamiento y almacenamiento, son ideales para esta tarea, ya que manejan y analizan grandes cantidades de datos de manera rápida y eficiente. Además, Huawei ofrece soluciones de IA y aprendizaje automático para analizar y descubrir patrones moleculares en *M. tuberculosis*. Por lo que la capacidad de procesamiento y análisis de grandes cantidades de datos en plataformas de Huawei son clave para el desarrollo del proyecto de investigación de TB-DR.

Ventajas de las plataformas de Huawei

La alianza UNAM-Huawei ofrece acceso a recursos de vanguardia, como servidores Atlas 800 y Thaisan, así como a una amplia gama de capacitaciones y certificaciones en IA y computación avanzada. Estos recursos permitieron a nuestro equipo desarrollar y mejorar nuestro modelo de IA de manera eficiente y efectiva en comparación con otras convocatorias. Huawei cuenta con procesadores de alto rendimiento que procesan grandes cantidades de datos en tiempos reducidos (por ejemplo, procesadores Ascend-Huawei. Huawei también ofrece soluciones de almacenamiento de datos de alta capacidad (nube, dispositivos externos y servidores).

La colaboración Huawei-UNAM ha permitido una transferencia tecnológica efectiva desde la investigación académica hasta el ámbito clínico. Al utilizar los recursos lógicos, como ejemplos de código fuente y el foro de desarrolladores Huawei-Arcend, nuestro grupo de trabajo de la UABC ha podido intercambiar conocimientos y colaborar con expertos en IA y profesionales de la salud. Esta interacción ha facilitado la adaptación de nuestro modelo de IA a las necesidades específicas del entorno clínico y la implementación del modelo en la práctica médica.

Además, las capacitaciones y certificaciones ofrecidas por Huawei y la DGTIC han permitido a nuestro equipo adquirir habilidades y conocimientos esenciales en el desarrollo de modelos de IA, así como una comprensión más profunda de las aplicaciones clínicas. La

transferencia tecnológica efectiva entre la investigación académica y el ámbito clínico es crucial para el éxito de nuestro proyecto, ya que permite acelerar la adopción de nuestra solución de IA en la práctica médica y, en última instancia, mejorar la calidad de vida de los pacientes afectados por la TB-DR.

Hay varias razones por las cuales Huawei se destaca en comparación con otras plataformas:

1. Capacidad de procesamiento y almacenamiento.
2. Soluciones de IA y aprendizaje automático.
3. Experiencia en proyectos de investigación.
4. Recursos y capacitación.

Sobre cómo se integra el equipo de trabajo

La necesidad de atender los problemas de salud del estado de Baja California nos motivó a conformar un grupo multidisciplinario (bioingenieros, médicos, computólogos y epidemiología molecular), que desarrolla sus actividades en el Laboratorio de Bionanoingeniería de la Facultad de Ingeniería Arquitectura y Diseño y el Laboratorio de Epidemiología y Ecología Molecular de la Escuela de Ciencias de la Salud, ambos de la Universidad Autónoma de Baja California. Así como en la Clínica y Laboratorio de Tuberculosis del Hospital General de Tijuana.

La líder del proyecto tiene amplia experiencia en el área de la IA, dirigiendo proyectos de machine learning y deep learning con aplicación a problemas del área médica y de la salud. La Dra. Raquel Muñiz-Salazar cuenta con una larga trayectoria en el análisis epidemiológico y molecular de enfermedades infecciosas ha trabajado de la mano de varias instituciones de salud en el combate de la enfermedad de TB tanto en el diagnóstico como en clasificación de la farmacorresistencia. El grupo también cuenta con los bioingenieros Guillermo Paredes y Miguel Ángel Guerrero, con experiencia en procesamiento de datos biomédicos, tanto curación y gestión de las bases de datos. Además, se cuenta con la participación del médico Ricardo Perea, con experiencia en investigación biomédica en el área de farmacogenética y análisis de genomas bacterianos.

Nuestro equipo forma parte de una red de trabajo interdisciplinaria para el estudio de la tuberculosis (RE-MITB), por lo que los resultados y procedimientos son expuestos para colaboración con otros grupos de trabajo. Al mismo tiempo, se cuenta con vías de comunicación con departamentos de salud en el estado de Baja California con el fin de dar a conocer las actividades del grupo de trabajo y dar lugar a consolidación de actividades cooperativas.

Conclusiones

El estudio de la TB farmacorresistente contribuye al fortalecimiento de la investigación en IA en México al involucrar a científicos de diferentes disciplinas e instituciones, quienes colaboran y fortalecen el proyecto. Así también permite mejorar la atención médica en México al desarrollar nuevos tratamientos y estrategias para combatir la TB farmacorresistente, impactando de manera positiva en la disminución de las tasas de mortalidad y mejorando la calidad de vida de las personas afectadas por esta enfermedad.

Lo que permitirá una mayor visibilidad de los proyectos de IA por parte de las instituciones públicas y los sistemas de salud al brindarles una herramienta que permite impactar de manera positiva en las problemáticas que presenta el país.

Con el apoyo de esta convocatoria, se brindó la oportunidad de adquirir habilidades para procesar la información de datos genómicos, el uso de espacio y poder de cómputo para adaptar esta información y generar información que apoye a la toma de decisiones clínicas. Tam-

bién, el proyecto presentó una gran oportunidad de llevar estos modelos desarrollados en un ambiente de investigación del área computacional a una implementación en campo clínico. Estos primeros resultados fomentarán el interés por las actividades de investigación y formación académica interdisciplinaria, así como oportunidades de formar y capacitar a los integrantes de instituciones públicas en la utilización de estas tecnologías para la toma de decisiones. Nuestro proyecto tiene un impacto significativo en el desarrollo de la IA en México, ya que aborda un problema de salud pública importante, como es la tuberculosis farmacorresistente, utilizando técnicas avanzadas de IA y la infraestructura de la alianza UNAM-Huawei. Nuestro equipo ha sido capaz de desarrollar un modelo innovador que puede mejorar la calidad de vida de los pacientes y contribuir al conocimiento científico en el campo de la IA. Además, nuestra colaboración con la UNAM y Huawei fomenta la creación de redes y la transferencia de conocimientos entre académicos, investigadores e industria, impulsando el crecimiento y la innovación en el ámbito de la IA en México.

BIBLIOGRAFÍA

- [1] "Global tuberculosis report", World Health Organization, Ginebra, 2022.
- [2] R. Muñiz Salazar and R. Laniado-Laborín, Eds., La Tuberculosis en México. Baja California, México: Universidad Autónoma de Baja California, México, 2018, p. 352.
- [3] World Health Organization, Manual operativo de la OMS sobre la tuberculosis. Módulo 3: Diagnóstico. Métodos de diagnóstico rápido para detectar la tuberculosis, 2020. Washington, D.C.: Organización Panamericana de la Salud; 2022. Licencia: CC BY-NC-SA 3.0 IGO. (2022) [Online]. Disponible en: <https://doi.org/10.37774/9789275325377>
- [4] C. J. Haug and J. M. Drazen, "Artificial Intelligence and Machine Learning in Clinical Medicine, 2023", en New England Journal of Medicine, 2023, vol. 388, pp. 1201-1208.
- [5] S. Kulkarni, N. Seneviratne, M. S. Baig, and A. H. A. Khan, "Artificial Intelligence in Medicine: Where Are We Now?", en Acad Radiol, Jan 2020, vol. 27, pp. 62-70.
- [6] World Health Organization, WHO Consolidated Guidelines on Tuberculosis, Module 4: Treatment - Drug-Resistant Tuberculosis Treatment. Geneva: World Health Organization, 2020.
- [7] InDRE and Secretaría de Salud, Lineamientos para la Vigilancia por Laboratorio de la Tuberculosis, (2019). Ciudad de México.
- [8] S. E. Sandoval-Azuara, R. Muñiz-Salazar, R. Perea-Jacobo, S. Robbe-Austerman, A. Perera-Ortiz, G. López-Valencia, et al., "Whole genome sequencing of Mycobacterium bovis to obtain molecular fingerprints in human and cattle isolates from Baja California, Mexico", en International Journal of Infectious Diseases, 2017, vol. 63, pp. 48-56.

- [9] A. P. Ortiz, C. Perea, E. Davalos, E. F. Velázquez, K. S. González, E. R. Camacho, et al., “Whole Genome Sequencing Links Mycobacterium bovis From Cattle, Cheese and Humans in Baja California, Mexico”, en *Frontiers in Veterinary Science*, August-03-2021, vol. 8.
- [10] A. C. Jiménez-Ruano, C. F. Madrazo-Moya, I. Cancino-Muñoz, P. M. Mejía-Ponce, C. Licon-Cassani, I. Comas, et al., “Whole genomic sequencing based genotyping reveals a specific X3 sublineage restricted to Mexico and related with multidrug resistance”, en *Scientific Reports*, 2021, vol. 11, p. 1870.
- [11] C. F. Madrazo-Moyza, I. Cancino-Muñoz, B. Cuevas-Cordoba, V. Gonzalez-Covarrubias, M. Barbosa-Amezcuca, X. Soberon, et al., “Whole genomic sequencing as a tool for diagnosis of drug and multidrug-resistance tuberculosis in an endemic region in Mexico”, en *bioRxiv*, 2019, p. 551481.
- [12] C. A. Perea Razo, E. Rodríguez Hernández, S. I. R. Ponce, F. Milián Suazo, S. Robbe-Austerman, T. Stuber, et al., “Molecular epidemiology of cattle tuberculosis in Mexico through whole-genome sequencing and spoligotyping”, en *PLoS One*, 2018, vol. 13, p. e0201981.
- [13] L. A. Rendón-Bautista, I. Álvarez-Maya, M. Sandoval-Díaz, J. C. Villanueva-Arias, N. Ayala-Chavira, and R. Zenteno-Cuevas, “Characterization of genetic diversity and clonal complexes by whole genome sequencing of Mycobacterium tuberculosis isolates from Jalisco, Mexico”, en *Tuberculosis*, Jul 2021, vol. 129, p. 102106.
- [14] D. A. V. Escárcega, C. A. P. Razo, S. G. Ruíz, S. L. S. Gallegos, F. M. Suazo, and G. J. C. Alarcón, “Analysis of Bovine Tuberculosis Transmission in Jalisco, Mexico through Whole-genome Sequencing”, en *J Vet Res*, Mar 2020, vol. 64, pp. 51-61.
- [15] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, “Machine learning in medicine: a practical introduction”, en *BMC Med Res Methodology*, Mar 19 2019, vol. 19, p. 64.
- [16] M. M. Ahsan, S. A. Luna, and Z. Siddique, “Machine-Learning-Based Disease Diagnosis: A Comprehensive Review”, en *Healthcare*, 2022, vol. 10, p. 541.
- [17] World Health Organization, Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance. Geneva: Licence: CC BY-NC-SA 3.0 IGO, 2021.
- [18] O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Cham (CH): Springer, 2022.
- [19] J. A. Grobler, A. S. Anderson, P. Fernandes, M. S. Diamond, C. M. Colvis, J. P. Menetski, et al., “Accelerated Preclinical Paths to Support Rapid Development of COVID-19 Therapeutics”, en *Cell Host Microbe*, Nov 11 2020, vol. 28, pp. 638-645.
- [20] J. Libiseller-Egger, J. Phelan, S. Campino, F. Mohareb, and T. G. Clark, “Robust detection of point mutations involved in multidrug-resistant Mycobacterium tuberculosis in the presence of co-occurrent resistance markers”, en *PLOS Computational Biology*, 2020, vol. 16, p. e1008518.

Fecha de recepción: julio 2023

Fecha de publicación: noviembre 2023

Cómo se cita

Dora-Luz Flores, Ricardo Perea-Jacobo, Guillermo Paredes, Miguel Guerrero, Raquel Muñiz Salazar “Combatiendo la tuberculosis farmacorresistente desde la Inteligencia Artificial” *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 8, noviembre, 2023. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

PLATAFORMA DE IA UNAM-HUAWEI PARA EL MAPEO DE INUNDACIONES MEDIANTE IMÁGENES SATELITALES

DOI: 10.22201/dgtic.26832968e.2023.8.5

Fernando Pech May (fernando.pech@cinvestav.mx)
*Tecnológico Nacional de México, Campus de los Ríos,
Balancán, Tabasco, México.*
ORCID: 0000-0002-8493-4534

Jorge Magaña Govea (totosaus@hotmail.com)
*Tecnológico Nacional de México, Campus de los Ríos,
División Académica de Ingeniería en Sistemas
Computacionales. Balancán, Tabasco, México.*
ORCID: 0009-0006-5522-1216

www.ties.unam.mx

Luis Antonio López Gómez (profesoritsr@gmail.com)
*Tecnológico Nacional de México, Campus de los Ríos,
Balancán, Tabasco, México.*

Germán Ríos Toledo (german.rt@tuxtla.tecnm.mx)
*Tecnológico Nacional de México, Campus Tuxtla Gutiérrez,
Departamento de Sistemas y Computación. Tuxtla Gutiérrez,
Chiapas, México.*

Fecha de recepción: Julio 2023 • Fecha de publicación: noviembre, 2023

Noviembre 2023 | número de revista 8 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

PLATAFORMA DE IA UNAM-HUAWEI PARA EL MAPEO DE INUNDACIONES MEDIANTE IMÁGENES SATELITALES

Resumen

El presente artículo aborda las vivencias del proyecto “Mapeo de inundaciones del sureste mexicano mediante imágenes satelitales Sentinel y aprendizaje Profundo” presentado en la convocatoria alianza UNAM-HUAWEI. En este se propuso desarrollar una estrategia para la detección y mapeo de inundaciones y zonas de riesgo, en la zona de los Ríos del estado de Tabasco, con el uso de imágenes satelitales Sentinel 1, Sentinel 2 y Aprendizaje profundo. El proyecto es realizado con la participación de profesores investigadores y alumnos, en su desarrollo se crearon muestras de entrenamiento, ejecutaron algoritmos de aprendizaje profundo haciendo uso de equipo de cómputo de alto procesamiento, accedidos a partir del programa UNAM-HUAWEI, herramientas de IA y conocimientos adquiridos como producto de las capacitaciones en la Plataforma HUAWEI y se logró la validación y optimización del modelo para la clasificación de zonas con presencia de inundaciones sobre las imágenes satelitales SAR de entrada.

Palabras clave:

Inteligencia Artificial, Aprendizaje profundo, Algoritmos de aprendizaje, Plataforma Huawei, UNAM-Huawei.

UNAM-HUAWEI AI PLATFORM FOR FLOOD MAPPING USING SATELLITE IMAGES

Abstract

This paper addresses the experiences of the project “Mapping floods in the Mexican south-east using Sentinel satellite images and deep learning” presented in the UNAM-HUAWEI alliance call. In this, it was proposed to develop a strategy for the detection and mapping of floods and risk areas, in the area of Los Ríos in the state of Tabasco, with the use of Sentinel 1, Sentinel 2 satellite images and Deep Learning. The project is carried out with the participation of research professors and students, in its development training samples were created, deep learning algorithms were executed using high-processing computing equipment, accessed from the UNAM-HUAWEI program, AI tools and knowledge acquired as a result of the training in the HUAWEI Platform and the validation and optimization of the model for the classification of areas with the presence of floods on the input SAR satellite images was achieved.

Keywords:

Artificial intelligence, Deep Learning, Learning Algorithms, Huawei platform, UNAM-Huawei.

PLATAFORMA DE IA UNAM-HUAWEI PARA EL MAPEO DE INUNDACIONES MEDIANTE IMÁGENES SATELITALES

Introducción

Los programas Nacionales Estratégicos del Gobierno Federal (PRONACES) tienen entre sus metas dar solución a problemas prioritarios nacionales para el bienestar social y el cuidado del medio ambiente [3]. Estos programas coinciden con los Objetivos de Desarrollo Sostenible (ODS) de la ONU [9], entre las que destacan: Salud, Agua, Soberanía alimentaria, energía y cambio climático, entre otros.

Para alcanzar estos objetivos es necesario el esfuerzo de investigadores, organizaciones sociales, desarrollo tecnológico, etc. La Ciencia de la Computación no está exenta de participar.

Los avances tecnológicos ofrecen un conjunto de herramientas que permiten realizar numerosas investigaciones dando solución a problemas relacionados con el medio ambiente, cambio climático, agricultura, inundaciones, etc.

Las inundaciones son consideradas entre los fenómenos naturales más destructivos, causando daños humanos y pérdidas económicas en el mundo [2]. En México, estados como Tabasco, sufren este fenómeno cada año, afectando las actividades económicas, salud humana, agricultura, ganadería entre otros.

En consecuencia, surge la necesidad de proponer estrategias para mitigar los daños ocasionados por las inundaciones, a través del uso de técnicas y herramientas de inteligencia artificial y percepción remota para el análisis de la cobertura terrestre.

En este contexto, propusimos desarrollar una estrategia para la detección y mapeo de inundaciones y zonas de riesgo, en la zona de los Ríos del estado de Tabasco, a través del uso de imágenes satelitales Sentinel 1 y Sentinel 2 y Aprendizaje profundo. La ejecución de algoritmos de aprendizaje profundo requiere recursos de cómputo de alto desempeño, que por su alto costo, son difíciles de adquirir. Por fortuna, existen programas, como el espacio de innovación UNAM-HUAWEI, que proporciona al investigador una plataforma completa para el acceso a recursos de cómputo de alto desempeño y desarrollar aplicaciones de inteligencia artificial para dar solución a problemas de carácter social.

En este artículo se describen las experiencias vividas en la creación del proyecto “Mapeo de inundaciones del sureste mexicano mediante imágenes satelitales Sentinel y aprendizaje Profundo” presentado en la alianza UNAM-HUAWEI.

Creación del Proyecto

De acuerdo con el Centro de Investigación sobre la epidemiología de las catástrofes (CRED, del inglés Centre for Research on the Epidemiology of Disasters) [2], en el 2021 se presentaron 432 catástrofes, de las cuales 223 fueron inundaciones.

Tabasco es una de las zonas más afectadas por las inundaciones. Sus características geomorfológicas dan lugar a un sistema fluvial dinámico y lluvias en todo el año. Posee una zona con grandes ríos (conformado por

los municipios de Balancán, Emiliano Zapata y Tenosique) propensos constantemente a inundaciones (ver Figura 1) [6]. La inundación más reciente tuvo lugar en el año 2020 causando severos daños a la propiedad, infraestructura, agricultura y ganadería [4].

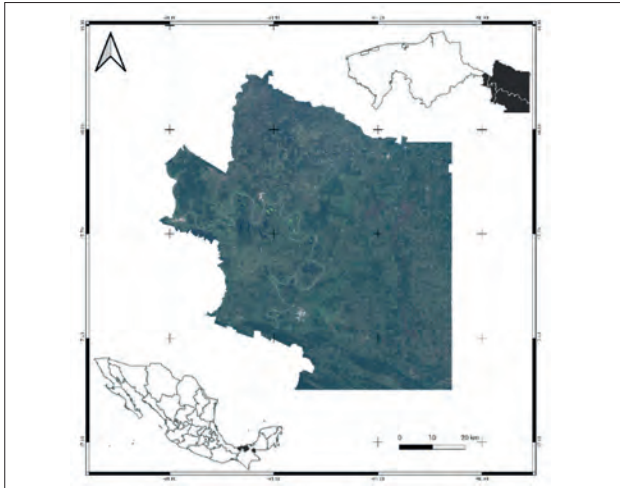


Figura 1. Ubicación geográfica de la zona de ríos del estado de Tabasco.

El auge de plataformas, herramientas satelitales y sensores, cada vez más avanzados, hacen posible recopilar una gran cantidad de datos de la tierra. Los tipos de imágenes que proporcionan los satélites pueden ser de tipo óptico (multiespectrales) o SAR (radar de apertura sintética) [1]. Estas imágenes han permitido realizar numerosas investigaciones enfocadas en la detección de inundaciones, análisis de suelo, monitoreo de daños naturales, etc [8]. Asimismo, las técnicas de aprendizaje profundo (DL), específicamente las Redes Neuronales Convolucionales (CNN), han emergido en los últimos años, utilizándose para el análisis de imágenes satelitales con resultados prometedores [10, 7].

Dada la problemática de la región, se propone una estrategia que mitigue el impacto de las inundaciones sobre la población y los sectores agrícola y ganadero del Estado de Tabasco mediante el uso de la inteligencia artificial e imágenes satelitales.

El grupo se centra en utilizar imágenes satelitales SAR y ópticas y redes neuronales convolucionales para el mapeo de inundaciones en la zona Ríos del estado de Tabasco, requiriéndose equipo de cómputo de alto rendimiento, para el manejo de grandes volúmenes de datos. En este sentido, el espacio y apoyo brindado por UNAM-HUAWEI ha sido fundamental.

A. Conformación del grupo de trabajo

El lugar donde se desarrolló el Proyecto fue el Instituto Tecnológico Superior de los Ríos (ITSR), ubicado en Balancán, Tabasco, zona geográficamente remota, atendiendo anualmente a estudiantes principalmente de Chiapas y Tabasco (ver Figura 2).

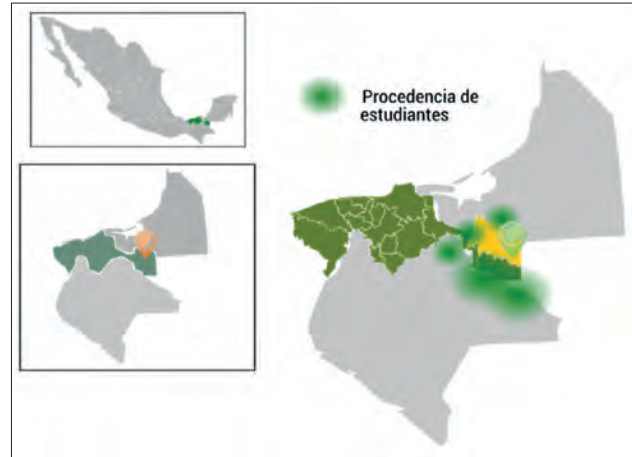


Fig 2. Ubicación del ITSR y procedencias de estudiantes que atiende.

El Cuerpo Académico (CA) “Cómputo Distribuido” del ITSR ha conformado un equipo de trabajo integrado por docentes investigadores y estudiantes involucrados en la realización de proyectos de investigación orientados a dar propuestas de solución a problemas regionales mediante el uso de la IA (ver Figura 3). Apoyados a través de la participación en convocatorias como el espacio de Innovación UNAM-HUAWEI para disponer de recursos de cómputo de alto desempeño carentes en instituciones como la nuestra.

B. Miembros del grupo de trabajo

Dr. Fernando Pech May. Doctor en Ciencias de la Computación con especialidad en Aprendizaje automático y web semántica. MATI. Jorge Magaña Govea. Maestro en Tecnologías de la información. Ing. Luis Antonio López Gómez. Investigadores de la academia de Ingeniería en Sistemas Computacionales del Tecnológico Nacional de México campus de los Ríos

Dr. Germán Ríos Toledo. Doctor en Ciencias de la Computación con especialidad del Procesamiento de Lenguaje Natural y aprendizaje automático. Investigador de la academia de Ingeniería en Sistemas Computacionales del Tecnológico Nacional de México campus Tuxtla Gutiérrez.

La participación de estudiantes ha sido fundamental en el proyecto. En este sentido, se invitó a estudiantes entusiastas y talentosos a participar en el proyecto de investigación. Participaron estudiantes de la carrera de Ingeniería en Sistemas Computacionales, 5 de servicio social y 4 estudiantes de residencia profesional con opción a titularse por tesis de investigación; todos matriculados a partir del quinto semestre de la carrera.

Con la integración de estudiantes en el proyecto, se busca contribuir en la formación integral de los futuros profesionales por medio del acercamiento al saber investigativo, centrados en su crecimiento profesional y humanístico, motivados por el desarrollo de su municipio, de su estado y de su país.



Fig. 3. Proceso de selección y formación de jóvenes investigadores a través de la convocatoria UNAM-HUAWEI.

Plataforma UNAM-HUAWEI para análisis de cuerpos de imágenes satelitales

La convocatoria estuvo dirigida al apoyo de proyectos sociales utilizando el aprendizaje profundo. La convocatoria provee recursos de hardware suficientes para el desarrollo de aplicaciones relacionadas con la Inteligencia Artificial. Facilitando con esta plataforma el análisis y procesamiento de datos masivos a través de la nube.

- 1. Recursos de cómputo de alto rendimiento.** Huawei posee plataformas de cómputo con procesadores para Inteligencia Artificial, lo que permite obtener alto rendimiento en el entrenamiento de algoritmos de aprendizaje profundo.
- 2. Herramientas y certificaciones para fortalecer al grupo de trabajo.** Huawei ofrece una plataforma de desarrollo y de implementación de modelos

de IA. Asimismo, una serie de librerías y códigos fuentes que pueden ser adaptados a cada proyecto. Asimismo, cursos y/o certificaciones relacionadas con aprendizaje automático, aprendizaje profundo, internet de las cosas, análisis de imágenes, entre otros. Lográndose con estas fortalecer al equipo de trabajo.

- 3. Capacitación para entender el impacto socio-ambiental del proyecto.** Identificación de los Objetivos de Desarrollo Sostenible (ODS), metas e indicadores relacionados a nuestro proyecto, así como los actores relacionados al mismo.
- 4. Capacitación para el uso de la plataforma HUAWEI IA.** A través de esta se logró la capacitación del grupo de investigación en la plataforma HUAWEI, con una arquitectura que ofrece soluciones de infraestructura con IA.

Con la ayuda de estos recursos de hardware, ha sido posible utilizar redes neuronales convolucionales facilitándose el manejo de imágenes a gran escala y complejidad con el uso de los equipos Ascend de Huawei.

Por otro lado, es importante destacar el uso de nuevos marcos de herramientas, como MidSpore, que facilitan el desarrollo de aplicaciones de IA. Esto ha permitido una reducción y facilidad en su desarrollo, ejecución eficiente, y la posibilidad de adaptabilidad en cualquier escenario de nuestro proyecto. Asimismo, el ahorro de tiempo para la creación de los modelos de red neuronal y su compatibilidad con distintas herramientas como PyTorch, Tensorflow, etc., pero con el aprovechamiento máximo de la potencia de cálculo de los procesadores Ascende de IA.

Entre los procesos realizados, a través de la plataforma Huawei, para la detección de cuerpos de agua destacan:

- Creación de muestras de entrenamiento.** Se utilizaron imágenes tipo SAR (radar de apertura sintética) de inundaciones en el estado de Tabasco de años anteriores. Se realizó la recolección de muestras obteniendo un conjunto de 1036 muestras repartidas en los diferentes escenarios utilizados en las imágenes satelitales. Una vez establecidas las muestras, se debe proceder a la exportación de los datos de entrenamiento.
- Entrenamiento del modelo.** Se genera un modelo basado en Aprendizaje Profundo usando como datos de entrada la colección de muestras que fueron

dispuestas para el proceso de entrenamiento. En la fase de aprendizaje del modelo propuesto en este trabajo, se utilizaron redes neuronales convolucionales del tipo U-Net [5].

- **Validación y optimización del modelo.** Se realizó la ejecución del modelo entrenado en el paso anterior para la clasificación de zonas con presencia de inundaciones sobre las imágenes satelitales SAR de entrada.

En la Figura 4 se muestra un ejemplo de procesamiento de imágenes SAR. En la Figura 5 se aprecia una muestra del mapeo de cuerpos de agua e inundaciones.

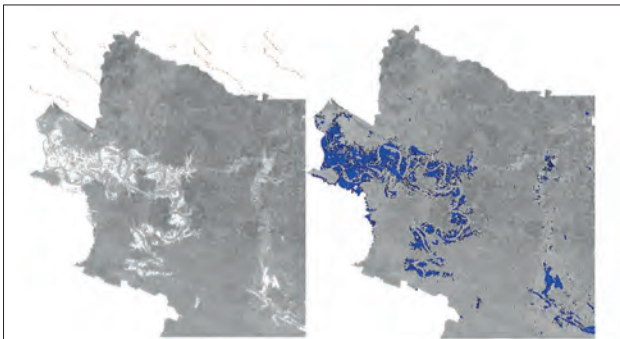


Fig. 4. Procesamiento de imágenes SAR utilizado en el proyecto de la convocatoria UNAM-HUAWEI. Izquierda: capa de diferencia, las áreas brillantes indican un cambio alto, las áreas oscuras poco cambio. Derecha: capa de extensión de inundación resultante aplicando un umbral de 1,25.

Oportunidades de la plataforma UNAM-HUAWEI

El espacio de innovación UNAM-HUAWEI ofrece una plataforma que permite realizar investigación y desarrollo tecnológico en el área de inteligencia artificial, aprendizaje automático, y en la ciencia de datos. Las herramientas que proporciona hacen posible impulsar proyectos de impacto social para dar solución a problemas regionales. Las unidades de procesamiento gráfico de los servidores son capaces de analizar y procesar videos e imágenes de alta definición útiles en el apren-

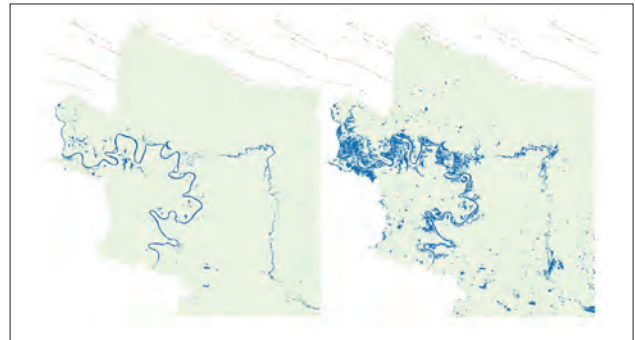


Fig. 5. Imágenes de muestra del mapeo e identificación de cuerpos de agua e inundación en la región Ríos de Tabasco en el proyecto de la convocatoria UNAM-HUAWEI. Izquierda: cuerpos de agua en temporada seca 2020. Derecha: cuerpos de agua y áreas inundadas.

dizaje automático. Asimismo, un conjunto de talleres y certificaciones para el manejo del hardware, plataforma para el procesamiento, redes neuronales profundas y almacenamiento de datos.

Esto hace posible formar recursos humanos expertos en inteligencia artificial y promover el uso de la Inteligencia Artificial para abonar en los programas nacionales estratégicos del país.

Conclusiones

El espacio de innovación UNAM-HUAWEI ha sido fundamental para el desarrollo del proyecto de mapeo de inundaciones. Su gran capacidad de procesamiento ha hecho posible crear, validar y optimizar el modelo para clasificar la presencia de inundaciones en la zona de estudio. Los códigos fuentes de acceso libre proporcionado por la plataforma HUAWEI permitieron reducir el tiempo de desarrollo. Por otro lado, el soporte proporcionado por expertos de la UNAM, permitió mejorar el proyecto para una segunda convocatoria. Con este tipo de espacios de innovación es posible sentar las bases en el uso de la inteligencia artificial para recopilación y manejo de datos, mejora de modelado a través de escalas espaciotemporales.

BIBLIOGRAFÍA

- [1] A. Shakya, M. Biswas, and M. Pal, “Fusion and Classification of SAR and Optical Data Using Multi-Image Color Components with Differential Gradients,” *Remote Sens.*, vol. 15, no. 1, p. 274, Jan. 2023. <https://doi.org/10.3390/rs15010274>
- [2] CRED, “2021 Disasters in Numbers,” <https://reliefweb.int/report/world/2021-disasters-numbers>, Accessed: Nov. 20, 2022. [acceso: 2022-11-20].
- [3] Conacyt, “Programas Nacionales Estratégicos,” [Online]. Disponible en: <https://conacyt.mx/pronaces/>
- [4] E. Cuevas & R. Norton, “Inundaciones de 2020 en Tabasco: Aprender del pasado para preparar el futuro,” Technical Report, Centro Nacional de Prevención de Desastres, Ciudad de México, 2022.
- [5] G. Konapala, S. Kumar, & A. Khalique, “Exploring Sentinel-1 and Sentinel-2 diversity for flood inundation mapping using deep learning,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 180, pp. 163–173, 2021.
- [6] J. Paz, F. Jiménez, & B. Sánchez, “Urge manejo del agua en Tabasco,” Technical Report, Universidad Nacional Autónoma de México y Asociación Mexicana de Ciencias para el Desarrollo Regional A.C., Ciudad de México, 2018.
- [7] L. Alzubaidi, J. Zhang, A. Humaidi, & A. Al-Dujaili, “A Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions,” *Journal of Big Data*, vol. 53, no. 8, 2021.
- [8] M. Chini, R. Pelich, L. Pulvirenti, N. Pierdicca, R. Hostache & P. Matgen, “Sentinel-1 insar coherence to detect flood-water in urban areas: Houston and hurricane harvey as a test case”. *Remote Sensing*, Vol. 11, 2022.
- [9] UN General Assembly, “Transforming our world: the 2030 Agenda for Sustainable Development”, 21 October 2015, A/RES/70/1, available at: <https://www.refworld.org/docid/57b6e3e4.html> [acceso 20 February 2023].
- [10] Y. Bengio, A. Courville & P. Vincent, “Representation learning: A review and new perspectives”. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, 2013.

Fecha de recepción: julio 2023

Fecha de publicación: noviembre 2023

Cómo se cita

Fernando Pech May, Jorge Magaña Govea, Luis Antonio López Gómez y Germán Ríos Toledo, “Plataforma de IA UNAM-HUAWEI para el mapeo de inundaciones mediante imágenes satelitales” *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 8, noviembre, 2023. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

ROBOT AUTÓNOMO PARA LA DETECCIÓN Y CLASIFICACIÓN DE FRESAS EN CAMPOS DE CULTIVO MEDIANTE DEEP LEARNING

DOI: 10.22201/dgtic.26832968e.2023.8.6

Luis Andrés Moreno Jiménez (luismj@cio.mx)
*Universidad Politécnica de Guanajuato, Centro de
Investigaciones en Óptica en el Laboratorio de Percepción y
Robótica. León, Guanajuato, México.*
ORCID: 0009-0000-9155-1462

Gesem Gudiño Mejía (gmejia@cio.mx)
Intel Labs
ORCID: 0000-0003-1634-7239

Andrés Montes de Oca Rebolledo (andresmr@cio.mx)
*Universidad de California Davis, Depto. Ingeniería Biológica y Agrícola.
California, Estados Unidos.*
ORCID: 0000-0002-6639-647X

Gerardo Flores (gflores@cio.mx)
*Universidad Politécnica de Guanajuato, Centro de Investigaciones en Óptica
en el Laboratorio de Percepción y Robótica. León, Guanajuato, México.*
ORCID: 0000-0003-2107-7405



www.ties.unam.mx

Fecha de recepción: julio de 2023 • Fecha de publicación: noviembre, 2023

Noviembre 2023 | número de revista 8 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

ROBOT AUTÓNOMO PARA LA DETECCIÓN Y CLASIFICACIÓN DE FRESAS EN CAMPOS DE CULTIVO MEDIANTE DEEP LEARNING

Resumen

Se realiza un acercamiento a las principales tecnologías para el desarrollo de aplicaciones móviles (*apps*) disponibles en el mercado actual. Se explican las principales características de las tecnologías: web móvil, de código nativo tanto tradicionales como híbridas, así como las ventajas y desventajas de utilizar cada una de ellas. Se abordan los problemas que los desarrolladores experimentan al emplear plataformas híbridas, se brindan sugerencias para la elección de la mejor tecnología y se ofrecen algunas conclusiones.

Palabras clave:

Agricultura, Deep Learning, Detección, Robot autónomo, Tecnología.

AUTONOMOUS ROBOT FOR STRAWBERRY DETECTION AND CLASSIFICATION IN CROP FIELDS THROUGH DEEP LEARNING

Abstract

The development of autonomous systems with applications in agriculture contributes significantly to achieve greater efficiency in daily tasks, such as monitoring production in the agricultural field. This work addresses precisely this issue through the proposal of an autonomous robot, which uses Deep Learning algorithms to detect and classify strawberries automatically according to their ripeness stage. This technological application aims to provide the farmer with precise information on the amount of strawberries present in a crop, as well as their state of maturity and exact location, which provides the farmer with tools to plan an efficient harvest. In real strawberry growing scenarios, there are factors that affect the accuracy of strawberry detection. However, given the results obtained in the field test, this system proves to be accurate in this task.

Keywords:

Agriculture, Deep Learning, Detection, Autonomous robot, Technology.

ROBOT AUTÓNOMO PARA LA DETECCIÓN Y CLASIFICACIÓN DE FRESAS EN CAMPOS DE CULTIVO MEDIANTE DEEP LEARNING

Introducción

La agricultura es una actividad muy importante para la humanidad [1], y a medida que la población siga creciendo, la innovación y el uso de métodos eficientes serán cada vez más importantes para aumentar la cantidad y calidad de la producción de alimentos. Uno de los retos de la agricultura es la recolección oportuna de frutos, una tarea delicada, ya que existe una ventana muy corta de tiempo de cosecha. A la vez, esta tarea también es tardada y requiere de arduo trabajo, especialmente en campos de cultivos grandes. Afortunadamente, los avances en la implementación de inteligencia artificial en robots autónomos están haciendo que estas tareas sean más eficientes [2], [3], [4] al permitir realizar la clasificación por nivel de madurez de forma autónoma.

En la actualidad, el aprendizaje profundo o Deep Learning [5], ha permitido desarrollar algoritmos capaces de imitar la manera en la que el cerebro humano procesa la información, utilizando redes neuronales artificiales que pueden aprender y reconocer patrones complejos. Estas redes se entrenan con grandes cantidades de datos para identificar patrones y características comunes entre ellos.

Imagine un vehículo terrestre de 25 centímetros de altura recorriendo un campo de fresas en el estado de Guanajuato, México, captando imágenes del entorno, clasificando automáticamente las fresas por su grado de

madurez, y generando un mapa con su posición a medida que avanza. Aunque pueda parecer demasiado idealista, esto es exactamente lo que se ha conseguido en un reciente proyecto de investigación [6]. Se ha desarrollado un sistema de monitoreo de fresas implementando un algoritmo de Deep Learning en un robot autónomo, lo que representa una solución innovadora y eficiente para la agricultura.

En este artículo, exploramos el desarrollo e implementación de esta tecnología en la tarea de detección y clasificación de fresas en la agricultura, así como los detalles técnicos del desarrollo del robot autónomo. A través de la descripción del proceso de integración del algoritmo de clasificación, el proceso de detección, los componentes clave del robot autónomo, y los resultados obtenidos, se presentará una visión completa de cómo se logró este avance tecnológico.

Alianza Tecnológica: LAPyR y Huawei-UNAM

El Laboratorio de Percepción y Robótica (LAPyR), fundado en 2017 por el Dr. Gerardo Flores, se ha destacado por su labor en la formación de investigadores en los campos de control, robótica e inteligencia artificial. El LAPyR fomenta la innovación a través de la sinergia entre teoría y práctica, centrándose en la resolución de problemas reales y en la creación de soluciones efectivas. Este laboratorio cuenta con un equipo interdisciplinario

especializado en áreas como Deep Learning aplicado a la percepción, agricultura de precisión, robótica móvil y visión artificial. Además, promueve la integración de estudiantes e investigadores de diferentes niveles académicos, desde universitarios hasta posdoctorados.

También mantiene colaboraciones activas a nivel nacional e internacional. En este contexto, se ha hecho una colaboración con Huawei-UNAM, enfocada en un innovador proyecto de monitoreo de fresas mediante algoritmos de Deep Learning y un robot autónomo. Huawei-UNAM ha apoyado el proyecto proporcionando recursos físicos (hardware), software (lógicos) y capacitación a los investigadores del LAPyR, además del acceso a los servidores avanzados de Huawei, foros para desarrolladores y transferencia tecnológica. A diferencia de otras convocatorias o colaboraciones, esta alianza ha proporcionado un apoyo integral y especializado en el ámbito de la inteligencia artificial. Esto ha permitido a los investigadores de LAPyR explorar y expandir sus conocimientos en el campo de la inteligencia artificial aplicada a la agricultura.

Evaluación de la madurez de las fresas y detección mediante inteligencia artificial

Estar en el campo rodeado de fresas de diferentes tamaños y colores e intentar cosechar las que están a punto de entrar a su madurez comercial es una tarea que requiere de mucho esfuerzo y dedicación, ya que si se cosechan demasiado verdes no madurarán y si están maduras llegarán a los centros de distribución en mal estado, por lo que hay que comprobar constantemente la madurez de la cosecha. Utilizando técnicas de procesamiento de imágenes se puede detectar la madurez de la fresa basándose en su cambio de color, de verde a rojo, como lo haría normalmente un agricultor.

¿Cómo se logra detectar este cambio en las imágenes? Mediante la transformación del espacio de color de una imagen, específicamente utilizando el espacio de color CIE Lab [7]. Este espacio de color cuenta con 3 canales, L (Luminosidad), a (tonos rojos y verdes), b (tonos azules y amarillos), que permiten percibir los cambios de color de manera uniforme. CIE Lab facilita la detección de diferencias entre el verde y el rojo en el canal a sin verse afectado por los cambios de iluminación.

El proceso de clasificación comienza con la imagen RGB original, con canales rojo R, verde G, y azul B, a la cual se le aplica la transformación al espacio de color CIE

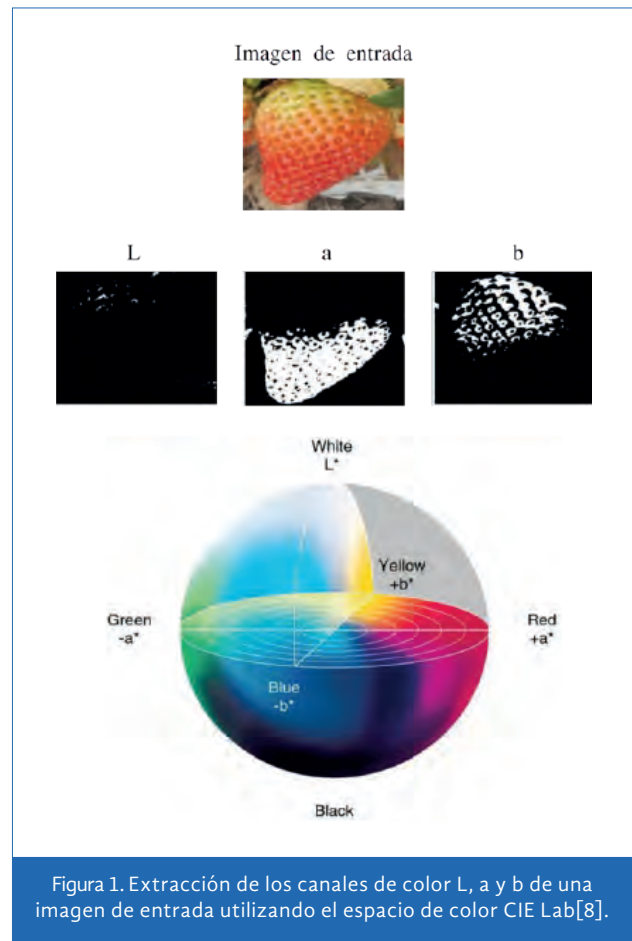


Figura 1. Extracción de los canales de color L, a y b de una imagen de entrada utilizando el espacio de color CIE Lab[8].

Lab. Posteriormente, a cada canal resultante (L, a, b) se le aplican límites para separar los tonos propios de la fresa del entorno. La Figura 1 muestra las tres capas del CIE Lab resultantes.

Tras esto se suman las capas (L + a + b), y se realizan algunas operaciones morfológicas para obtener una máscara con la forma aproximada de la fresa en la imagen original, sin perder información relevante de la misma, como forma y color. Ver figura 2.



Figura 2. Máscara obtenida del proceso de binarización, seguida de dilatación, erosión y aplicación de la máscara a la imagen RGB original.

Una vez extraída la segmentación de la fresa en formato RGB, se utiliza el canal rojo R de la imagen para la clasificación de madurez, ya que está fuertemente relacionado con el nivel de madurez de la fresa. El nivel de la madurez de la fruta se clasifica en tres niveles: fresa inmadura, semi-madura, y madura, según la intensidad media de sus píxeles. Lo anterior se ilustra en la Figura 3. Si bien las técnicas de procesamiento de imágenes son muy útiles para clasificar la madurez de las fresas, tienen sus limitaciones, como la presencia de objetos diferentes a la fresa que compartan los mismos colores, lo que puede generar falsos positivos y que dichos objetos sean identificados como fresas sin serlo. Para solucionar este problema y mejorar la precisión en la detección de fresas, se aplica Deep Learning, específicamente utilizando redes neuronales convolucionales (CNN) para la identificación de las fresas por forma y textura.

Las CNN, como señala LeCun, uno de los pioneros de la inteligencia artificial, son un tipo de red neuronal, diseñada para trabajar con imágenes, que utiliza convoluciones en lugar de multiplicaciones matriciales para aprovechar los datos de entrada [9]. En lugar de mirar toda la imagen y tratar de encontrar patrones, la CNN divide la imagen en diferentes partes y las procesa por separado, buscando patrones en los datos y extrayen-

do características generales como líneas o bordes, para luego utilizar estos patrones para formar objetos más complejos, permitiendo detectar y clasificar con mucha más precisión los objetos de la imagen, lo que es muy útil para tareas de reconocimiento.

En una CNN, como se muestra en la Figura 4, las múltiples capas trabajan juntas para realizar la tarea de clasificación. La primera etapa es la extracción de características, en la cual la CNN divide la imagen en fragmentos y procesa cada uno por separado para buscar

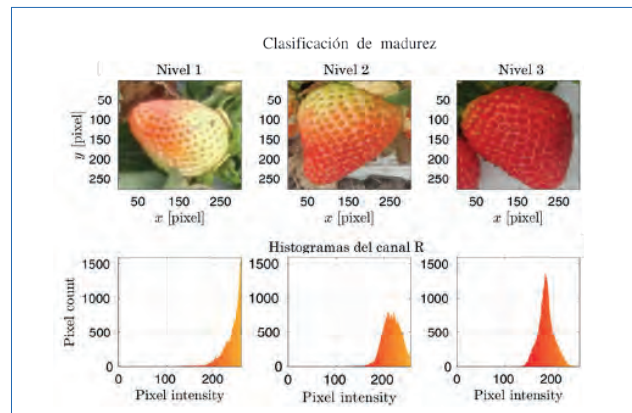


Figura 3. Histograma de color en el canal R de la imagen, clasificado en 3 niveles: inmaduro, semi-maduro y maduro.

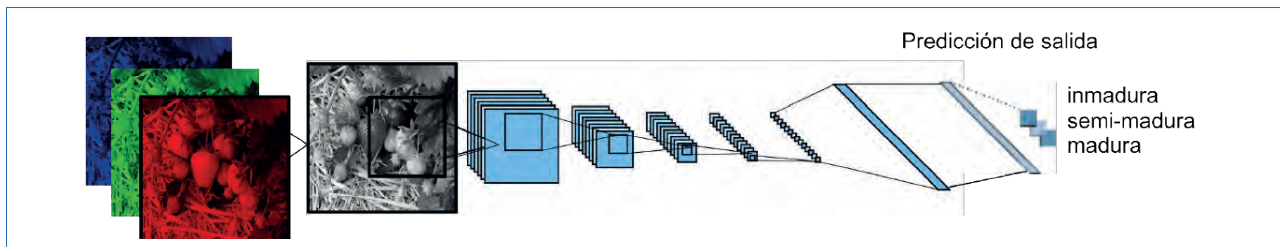


Figura 4. Arquitectura de la CNN utilizada en este trabajo.

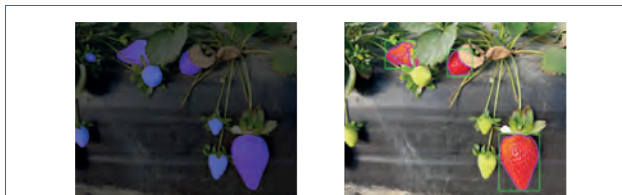


Figura 5. Detección de fresas en el conjunto de datos. Imagen izquierda: Muestra la imagen segmentada correspondiente a la entrada de la imagen. Imagen derecha: Muestra las fresas maduras detectadas por la CNN, representadas por cuadros verdes en las regiones de interés.

patrones en los datos. A medida que la red avanza, las dimensiones se reducen, encontrando características cada vez más complejas. Al final de la red hay neuronas simples que realizan la clasificación en una o más clases, según las características encontradas.

En el caso de la detección de fresas, la CNN se entrenó con imágenes de fresas del popular conjunto de datos ImageNet [10], y se aumentó con un conjunto de imágenes de fresas en el campo de prueba, lo que le permitió a la red identificar la forma y textura de las fresas para su posterior clasificación. En este proyecto

se contó con la colaboración de Huawei, que proporcionó recursos de cómputo de alto desempeño, como sus servidores avanzados, lo que permitió un mayor rendimiento en el procesamiento de información y en el entrenamiento de la IA.

El uso de los servidores avanzados de Huawei en este proyecto permitió acelerar significativamente el tiempo de entrenamiento de la CNN y mejorar su capacidad para trabajar con conjuntos de datos más grandes, lo que permitió a la red generalizar de mejor manera las características que describen a una fresa. Estos servidores de alto desempeño son especialmente útiles en el campo de la inteligencia artificial, ya que pueden manejar eficientemente la gran cantidad de datos y operaciones matemáticas necesarias para entrenar una red neuronal.

Integración de un robot autónomo

La columna vertebral de este proyecto es el vehículo robótico terrestre, ya que en él se monta y aplica el hardware y software desarrollado. Si bien las redes neuronales convolucionales son muy útiles, requieren de un alto nivel de procesamiento computacional y, por lo tanto, necesitan equipos de mayor tamaño y consumo de energía. Esto dificulta el uso de vehículos compactos que puedan trasladarse entre los surcos del campo. La solución a esto es el uso de computadoras



Figura 6. Plantío de fresas en Pénjamo, Guanajuato, México.

embebidas, en este caso la Jetson Xavier, que pueden procesar una gran cantidad de información gracias a su GPU integrada. Con esto, se tiene la capacidad de ejecutar algoritmos de Deep Learning sin necesidad de estar conectado a equipos externos. Además, su bajo consumo de energía facilita su uso con baterías, lo que la hace ideal para su integración en robots autónomos y en tareas de monitoreo.

El terreno en el que se trabajará, mostrado en la Figura 6, es representativo de la mayoría de los cultivos en México, con un suelo irregular y espacios estrechos. Por esta razón, se seleccionó un Trail Crawler modelo TRX-4 de la marca Traxxas® adaptado para convertirse en un vehículo tipo rover.

Una vez que se tiene el vehículo adecuado, es necesario adaptarlo para que se convierta en un robot autónomo, aquí es donde la instrumentación cobra importancia. Este vehículo se ha equipado con un controlador automático y otros sensores como GPS, una cámara estereoscópica ZED mini, ver Figura 7. Así, el rover puede navegar automáticamente mediante un programa de planificación de trayectoria o manualmente con un control remoto.

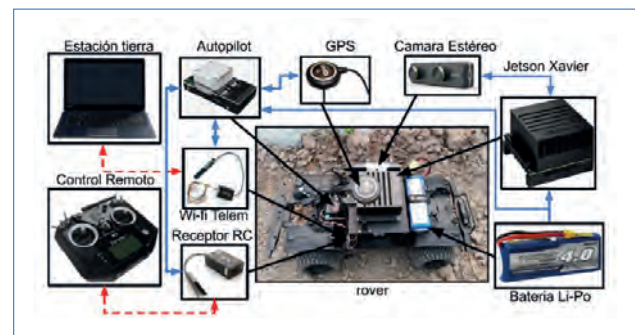


Figura 7. Componentes principales del vehículo.

Lo más importante es que estas características hacen que el rover sea ligero y compacto, lo que le permite navegar fácilmente por el campo de fresas, recolectando información importante y fiable de las fresas. Este robot podría ser una herramienta valiosa para la investigación agrícola y un ejemplo de cómo la tecnología puede contribuir al avance en este sector.

Implementación en el campo de fresas y resultados

Una vez completados los pasos para implementar la clasificación y la detección de las fresas por el robot autónomo, el funcionamiento del robot se realiza de acuerdo con lo mostrado en la Figura 8. En esta figura se observa

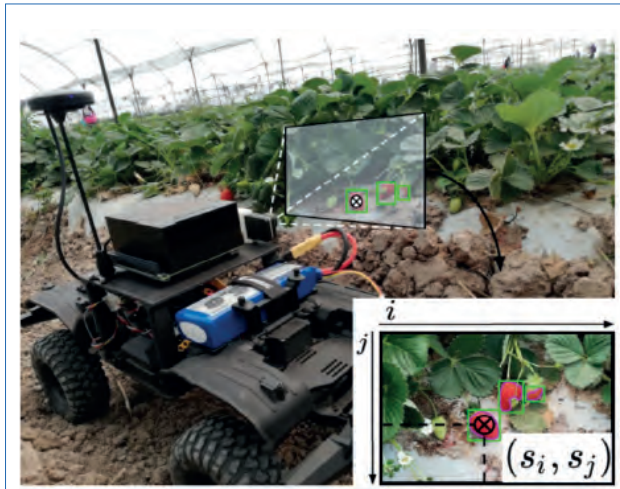


Figura 8. Rover detectando fresas en un entorno natural. En la esquina inferior derecha se puede ver la imagen capturada con las coordenadas del centroide de la fresa detectada (s_i, s_j).

cómo el rover navega paralelo a las fresas, y gracias a la cámara estereoscópica ubicada en su costado, a medida que avanza, el sistema identifica las fresas en su campo de visión, calcula su centroide y, mediante transformaciones geométricas, determina su posición local y global.

Sorprendentemente, el sistema logra detectar el 72% de las fresas totales en el campo y alcanza una precisión de detección del 93.75% en los escenarios más complejos como se ve en la Figura 9. Estos resultados superaron a los obtenidos por otros sistemas similares [11], [12]. Además, la colaboración con Huawei-UNAM generó valiosas oportunidades y lecciones aprendidas



Figura 9. Se muestra la detección de fresas maduras en el campo utilizando el rover. Se puede apreciar las fresas maduras, mientras que las fresas inmaduras se han ignorado.

en términos de capacidades tecnológicas y transferencia tecnológica. Un ejemplo concreto de esto es la generación de código en GitHub, disponible para su uso por otros investigadores, la industria agrícola y estudiantes interesados en el desarrollo de soluciones similares.

Adicionalmente, el sistema demostró ser eficiente en términos de velocidad de detección, procesando 23 imágenes por segundo, lo que demuestra que la integración de este tipo de robots autónomos es eficaz. Con esto se permite la navegación autónoma en el campo y la obtención de imágenes de alta calidad para su análisis posterior.

Por último, se calcula la posición de las fresas y se colocan en un mapa distinguiendo su nivel de maduración, facilitando al agricultor la identificación de su cosecha. La Figura 10 ilustra el proceso completo de este sistema: detección de la fresa mediante Deep Learning, extracción de información de intensidad para definir madurez, identificación de posición respecto al vehículo, su transformación a coordenadas globales y, finalmente, visualización de detecciones y ubicaciones en un mapa.

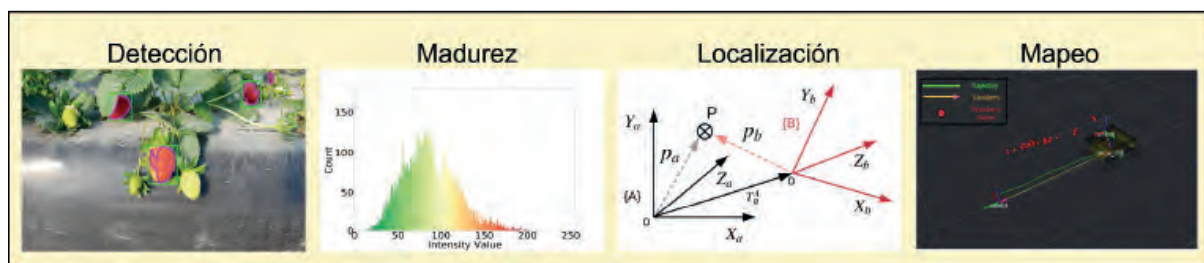


Figura 10. Se muestra una visión general del sistema completo propuesto, que consta de varios componentes clave.

Conclusión

La colaboración con Huawei-UNAM impulsó el desarrollo de este proyecto que combina tecnología avanzada, Deep Learning y cooperación empresarial-académica, logrando soluciones de inteligencia artificial aplicadas en la agricultura. El sistema desarrollado ha demostrado ser eficiente y preciso en la detección y clasificación

de fresas en el campo. A pesar de los desafíos, como la detección de fresas ocultas y la variedad de estados de maduración, este proyecto representa un avance significativo. Es necesario continuar investigando para aplicar este enfoque a otros cultivos, buscando mejorar la productividad y sostenibilidad agrícola en beneficio de una población en constante crecimiento.

BIBLIOGRAFÍA

- [1] M. A. Altieri, "La agricultura tradicional como legado agroecológico para la humanidad," *PH: Boletín del Instituto Andaluz del Patrimonio Histórico*, vol. 29, no. 104, pp. 180-197, 2021.
- [2] H. Bhardwaj et al., "Artificial intelligence and its applications in agriculture with the future of smart agriculture techniques," in *Artificial Intelligence and IoT-Based Technologies for Sustainable Farming and Smart Agriculture*, IGI Global, 2021, pp. 25-39.
- [3] S. T. Doddamani, S. Karadgi, and A. C. Giriapur, "Multi-Label Classification of Cotton Plant with Agriculture Mobile Robot," in *Data Intelligence and Cognitive Informatics: Proceedings of ICDICI 2021*, Singapore, Springer Nature Singapore, 2022, pp. 759-772.
- [4] A. Sharma et al., "Machine learning applications for precision agriculture: A comprehensive review," in *IEEE Access*, vol. 9, pp. 4843-4873, 2020.
- [5] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," in *Neurocomputing*, vol. 452, pp. 48-62, 2021.
- [6] G. Mejía, A. M. de Oca, and G. Flores, "Strawberry localization in a ridge planting with an autonomous rover," in *Engineering Applications of Artificial Intelligence*, vol. 119, pp. 105810, 2023.
- [7] B. Hill, T. Roger, and F. W. Vorhagen, "Comparative analysis of the quantization of color spaces on the basis of the CIELAB color-difference formula," in *ACM Transactions on Graphics (TOG)*, vol. 16, no. 2, pp. 109-154, 1997.
- [8] E. Agudo, P. J. Pardo, H. Sánchez, Á. L. Pérez, and M. I. Suero, "A low-cost real color picker based on Arduino," in *Sensors*, vol. 14, no. 7, pp. 11943-11956, 2014.
- [9] LeCun, Y. Bengio, and G. Hinton, "Deep learning," in *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.
- [10] Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 248-255.
- [11] T. Shao, Y. Wang, G. Xuan, Z. Gao, Z. Hu, C. Gao, and K. Wang, "Assessment of strawberry ripeness using hyperspectral imaging," in *Analytical Letters*, vol. 54, no. 10, pp. 1547-1560, 2020.
- [12] Liu and D. Xiao, "Recognition method of mature strawberry based on improved SSD deep convolutional neural network," in *Bio-inspired Computing: Theories and Applications: 14th International Conference, BIC-TA 2019, Zhengzhou, China, November 22-25, 2019, Revised Selected Papers, Part II*, vol. 14, pp. 271-281, Singapore, 2020.

Fecha de recepción: julio, 2023

Fecha de publicación: noviembre de 2023

Cómo se cita:

L. A. Moreno, G. Mejía, A. Montes de Oca, G. Flores, "Robot Autónomo para la detección y clasificación de fresas en campos de cultivo mediante Deep Learning," *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 8, noviembre, 2023. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].



TIES Revista de Tecnología e Innovación en Educación Superior

INTELIGENCIA ARTIFICIAL AL SERVICIO DE LOS PEQUEÑOS PRODUCTORES DE DÁTIL EN MÉXICO

DOI: 10.22201/dgtic.26832968e.2023.8.7

Juan Pablo García Vázquez (pablo.garcia@uabc.edu.mx)
*Universidad Autónoma de Baja California, Facultad de Ingeniería.
Mexicali, Baja California, México.*
ORCID: 0000-0003-1787-7223

Ricardo Salomón Torres (ricardo.salomon@ues.mx)
*Universidad Estatal de Sonora. San Luis Río Colorado, Sonora,
México.*
ORCID: 0000-0002-6486-2131

Dalila Blanca Pérez Pérez (dalila.perez9@uabc.edu.mx)
*Universidad Autónoma de Baja California, Facultad de Ingeniería.
Mexicali, Baja California, México.*
ORCID: 0000-0001-9361-0438

Héctor Zatarain Aceves (hector.zatarain@uabc.edu.mx)
*Universidad Autónoma de Baja California, Facultad de Ingeniería,
Arquitectura y Diseño. Mexicali, Baja California, México.*
ORCID: 0000-0003-2899-2198

Jesús Eduardo Soto Vega (soto.jesus@uabc.edu.mx)
*Universidad Autónoma de Baja California, Facultad de Ingeniería. Mexicali,
Baja California, México.*
ORCID: 0000-0001-8043-4762

Edwin R. García Curiel (edwin.garcia.curiel@uabc.edu.mx)
*Universidad Autónoma de Baja California, Facultad de Ingeniería. Mexicali,
Baja California, México.*
ORCID: 0000-0002-0172-8226

www.ties.unam.mx

Fecha de recepción: julio de 2023 • Fecha de publicación: noviembre, 2023

Noviembre 2023 | número de revista 8 • ISSN 2683-2968

Acervos Digitales, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

INTELIGENCIA ARTIFICIAL AL SERVICIO DE LOS PEQUEÑOS PRODUCTORES DE DÁTIL EN MÉXICO

Resumen

El cultivo de la palma datilera en México se ha incrementado durante los últimos años, convirtiéndose en el tercer mayor productor de la variedad Medjool a nivel mundial. Sin embargo, los pequeños productores se enfrentan a varios retos en la recolección, clasificación y el empaque de este fruto, debido a que son procesos que se realizan principalmente de manera manual. El principal problema se presenta en la clasificación, donde se busca una correcta selección de la fruta a través de la identificación de diversos parámetros de calidad, como el estado de madurez. Una inapropiada selección del fruto, puede ocasionar muchas desventajas para el agricultor en su comercialización, siendo un sistema de visión artificial una posible alternativa de solución. Para poder procesar un alto volumen de información, fue necesario apoyarse en la infraestructura Huawei, en este caso se utilizó la arquitectura DaVinci la cual es diseñada exclusivamente para cómputo de inteligencia artificial (AI). Las principales contribuciones del proyecto fueron el desarrollo de aplicaciones de AI para apoyar a los pequeños productores de dátil en sus actividades de selección y clasificación. Así como, la generación de recurso humano especializado en el área de inteligencia artificial.

Palabras clave:

dátil Medjool, nivel de madurez, inteligencia artificial, Mindspore, arquitectura Da Vinci.

ARTIFICIAL INTELLIGENCE AT THE SERVICE OF SMALL DATE PRODUCERS IN MEXICO

Abstract

Date palm cultivation in Mexico has increased in recent years, making it the third largest producer of the Medjool variety in the world. However, small growers face several challenges in the harvesting, sorting and packing of this fruit, since these processes are mainly carried out manually. The main problem occurs in grading the product, where a correct selection of the fruit is sought through the identification of various quality parameters, such as the state of maturity. Poor selection of fruit for commercialization can cause many problems for the farmer, which makes an artificial vision system a viable alternative solution. In order to process a high volume of information, it was necessary to rely on Huawei's infrastructure, the DaVinci architecture, designed exclusively for artificial intelligence (AI) computation. The main contributions of the project were the development of AI applications to support small date producers in their sorting and grading activities. As well as the generation of human resources specialized in the area of artificial intelligence.

Keywords:

Medjool date, ripeness level, artificial intelligence, Mindspore, Da Vinci architecture.

INTELIGENCIA ARTIFICIAL AL SERVICIO DE LOS PEQUEÑOS PRODUCTORES DE DÁTIL EN MÉXICO

1. Idea de proyecto

El proyecto surge a partir de la preocupación de la Universidad Autónoma de Baja California, por contribuir en la solución de problemáticas locales y regionales. Una estrategia del departamento de posgrado fue generar reuniones para fomentar la vinculación con diversos sectores (p. ej. turístico, industrial y agrícola). En estas reuniones se atendieron las necesidades y problemáticas de diversos sectores productivos.

Con respecto a los pequeños productores de palma datilera del Valle de Mexicali, se identificó que se enfrentan a varios retos en lo que respecta a la recolección, clasificación y empaque del dátil Medjool, derivado a que estas actividades se realizan principalmente de forma manual (ver Figura 1). El desarrollo de este trabajo implica el uso de mano de obra especializada,

alta rotación, largas jornadas laborales, realizar tareas repetitivas, lo cual provoca errores en la correcta inspección de los atributos de calidad de la fruta, como su color (nivel de madurez), tamaño y textura.

El proceso de cosecha del dátil Medjool, consiste en que los recolectores de fruta escalan a la palma, agitan el racimo para que los dátiles maduros se desprendan y caigan en los contenedores. Esta actividad puede provocar que la fruta madura sufra daños en su textura o que se cosechen frutos en etapas previas a la maduración (ver Figura 2). Después los dátiles se colocan en bandejas, donde se separan los inmaduros y se agruparán en otras bandejas para exponerlo a la luz solar, para que continúen con su proceso de deshidratación, hasta que alcancen su plena madurez. Una inapropiada selección



Figura 1. Trabajadores seleccionando de forma manual el dátil Medjool.



Figura 2. Máscara obtenida del proceso de binarización, seguida de dilatación, erosión y aplicación de la máscara a la imagen RGB original.

del fruto puede ocasionar muchas desventajas para el agricultor en su comercialización.

El fruto de menor tamaño o dañados, comúnmente son separados principalmente para elaborar subproductos de dátil y en menor medida para consumo animal. Finalmente, la clasificación por tamaño de los dátiles Medjool y que tienen el grado de madurez requerido, son empacados en diversas presentaciones de peso en cajas de cartón.

Existen diversos trabajos en la literatura que han propuesto tecnología para apoyar a los pequeños productores de dátil, en los procesos de clasificación y selección. Por ejemplo, en [1] se propuso un sistema de visión computacional que utiliza imágenes RGB con dátiles. A partir de estas imágenes, se extrae automáticamente características externas de calidad en este fruto. Basándose en las características extraídas, clasifica los frutos en tres categorías de calidad definidas por expertos, las cuales fueron utilizadas para entrenar a una red neuronal, obteniéndose una precisión del 80%. De manera similar, en [2] diseñaron e implementaron un sistema de visión artificial para la clasificación auto-

mática de dátiles. Este sistema ayudó a medir el tamaño de la fruta y la deslaminación de la piel. La prueba de producción ha demostrado una mejora en la precisión de la clasificación de aproximadamente un 10% con respecto a la clasificación manual. Asimismo, se redujeron los costos por concepto de mano de obra en casi un 75% (sólo se requirieron 15 trabajadores en lugar de 60, lo cual equivale entre 2,800 y 3,600 USD al día, de acuerdo con los salarios mínimos por hora de los trabajadores) y se acortó el tiempo de procesamiento para la clasificación

Determinar el estado de madurez del dátil Medjool mediante métodos tradicionales de procesamiento de imágenes y métodos de aprendizaje automático es complicado. Esto se debe a que estos métodos están entrenados para extraer características del fruto como su apariencia, color (asociadas a los estados de madurez), la forma y la textura [2,3]. Sin embargo, no existen estudios en los que se haya realizado una extracción de características o un modelo predictivo para la clasificación de dátil variedad Medjool, en los cuales se describa que se utilice un sistema de visión artificial. Además, los modelos recientes no pueden determinar la clasificación de Medjool porque este fruto se cosecha, se clasifica, se empaca y se consume en su fase de Tamar (ver Figura 3).



Figura 3. Dátil Medjool en diversos niveles de madurez (DPP días posteriores de la polinización).

2. Equipo de Trabajo

Para realizar el proyecto, se integró un equipo multidisciplinario de investigadores de la Universidad Autónoma de Baja California (UABC), Campus Mexicali y de la Universidad Estatal de Sonora (UES). El perfil de los integrantes es presentado en la Tabla 1.

Tabla 1. Investigadores y estudiantes involucrados en el proyecto

Nombre	Especialidad	Cuerpo Académico - Programa Educativo
Ricardo Salomón Torres	Doctor en Ciencias especializado en genómica, biotecnología y ciencia de datos, asimismo es reconocido por su trabajo de investigación en la palma datilera y el dátil variedad Medjool.	Biotecnología y Desarrollo Sustentable.
Juan Juan Pablo García Vazquez	Doctor en Ciencias con interés en la agricultura de precisión, inteligencia artificial, visión artificial, minería de datos, interacción humano computadora.	Tecnología para ambientes inteligentes - Licenciado en Sistemas Computacionales (LSC).

Nombre	Especialidad	Cuerpo Académico - Programa Educativo
Héctor Zatarain Aceves Jesús Eduardo Soto Vega Edwin R. García Curiel	Doctor en Ciencias de la computación con conocimientos en diversas áreas de la inteligencia artificial, tales como redes neuronales, lógica difusa, algoritmos bioinspirados, aprendizaje automático y profundo.	Licenciado en Sistemas Computacionales.
Dalila Blanca Pérez Pérez	Estudiante de Doctorado en ciencias quién cuenta con estudios de Ingeniero agrónomo con maestría en producción animal.	Programa de Maestría y Doctorado en Ciencias e Ingeniería (MyDCI).

3. Infraestructura utilizada para el proyecto

Existen una gran diversidad plataformas tecnológicas que permiten el desarrollo de proyectos de inteligencia artificial, por ejemplo, Google AI Platform, TensorFlow, Microsoft Azure, por mencionar algunas. Estas plataformas se caracterizan por permitir crear, implementar y administrar modelos computacionales con código abierto (p ej. Python). Así como por proveer servicios a desarrolladores poco experimentados con fines específicos, tales como procesamiento de lenguaje natural, visión artificial, análisis de datos. El cómputo ofrecido por estas plataformas comúnmente emplea CPU o GPU, los cuales tienen diferentes arquitecturas y se construyeron con propósitos diferentes. La CPU es adecuada para una amplia variedad de cargas de trabajo, especialmente aquellas en las que la latencia o el rendimiento por núcleo son importantes. La CPU es un potente motor de ejecución y centra su reducido número de núcleos en tareas individuales y en realizar las tareas rápidamente. Por tanto, resulta especialmente buena para tareas que van desde la computación en serie hasta la ejecución de bases de datos [4]. Mientras que, la GPU es un circuito integrado para aplicaciones específicas, por ejemplo, acelerar las tareas específicas de renderización en 3D. Recientemente, las GPU han evolucionado para convertirse también en procesadores paralelos de uso más general, por lo que manejan una gama cada vez más amplia de aplicaciones [4]. Sin embargo, ni la CPU ni la GPU fueron diseñadas para realizar cómputo de inteligencia artificial. Por tanto, pueden resultar ineficientes para el cómputo de IA que se basa comúnmente en redes neuronales convolucio-

nales (CNN) que internamente realizan operaciones masivas de multiplicación de matrices o vectores.

Huawei desarrolló la arquitectura Da Vinci específicamente para cómputo de inteligencia artificial. Esta arquitectura proporciona unidades de cómputo escalar, vectorial y cúbicas [5]; operaciones comúnmente realizadas por las redes neuronales convolucionales. La arquitectura utiliza los procesadores Ascend, por ejemplo, Ascend 910 que se encuentran en servidores que pueden ser usados para entrenar y evaluar modelos basados en redes neuronales convolucionales. Por esta razón participamos en la convocatoria Alianza Huawei-UNAM, ya que, a diferencia de otras convocatorias que ofrecen recursos económicos, en la convocatoria Alianza se otorgaban recursos físicos (hardware), concretamente acceso a horas de procesamiento para entrenamiento y evaluación de modelos en servidores Huawei con procesadores Ascend. Los equipos disponibles eran Atlas 800 (Modelo 9000), Huawei Atlas 800 (Modelo 3010) o Huawei Thetaishan (Modelo 2280). Además, ofrecían recursos lógicos (software) y capacitación en temas relacionados con la inteligencia artificial, aprendizaje automático y profundo, certificaciones de nube Huawei, tales como Huawei Certified ICT Associate (HCIA), Huawei Certified ICT Professional (HCIP), Huawei Certified ICT Expert (HCIE) y talleres de innovación centrada en las personas (Design Thinking).

4. Oportunidades y lecciones aprendidas

El participar en la convocatoria Alianza Huawei-UNAM nos dejó varios aprendizajes. En la investigación, aprendimos a elaborar propuestas de investigación utilizando

una metodología de innovación centrada en las personas. Aplicar la metodología nos ayudó a identificar el impacto socio-ambiental del proyecto, describir el problema central, determinar sus causas y objetivos, definir el segmento meta y actores clave, así como su nivel de involucramiento. Lo anterior, permitió generar un plan de investigación para identificar la naturaleza y contexto de la problemática que se pretende resolver. Así como articular las metas del proyecto con indicadores alineados a los objetivos de desarrollo sustentable y medios de verificación. Finalmente, aprendimos a plantear una propuesta de valor que pudiera ser transmitida a diversas audiencias.

En lo que respecta a capacitación consideramos que es necesario reforzar los conocimientos de los participantes con respecto a la propiedad intelectual. Con el objetivo de que los diseños, productos o desarrollos tecnológicos resultantes de los proyectos se registren ante el Instituto Mexicano de la Propiedad Industrial (IMPI).

También se desarrollaron capacidades digitales en los miembros del equipo, que reforzaron conocimientos de inteligencia artificial, aprendizaje automático y profundo. Además, se obtuvieron conocimientos técnicos sobre la arquitectura Da Vinci de los procesadores Ascend, la infraestructura de nube Huawei, el despliegue de aplicaciones en contenedores (Docker¹), y el marco de cómputo de inteligencia artificial Mindspore². Lo anterior, nos permitió identificar los siguientes nichos de oportunidad: (1) Implementar algoritmos de clasificación o detección de objetos no implementados aún en el marco de desarrollo Mindspore; y (2) Proponer nuevos algoritmos que aprovechen la arquitectura Da Vinci de los procesadores Ascend para realizar tareas de clasificación o detección de objetos.

Asimismo, se identifica como oportunidad mejorar el manejo de errores en el marco de desarrollo Mindspore, con esto nos referimos a que si un programador comete un error en el código de su aplicación, el marco debe ofrecer mensajes de error descriptivos que le ayuden a recuperarse del error con facilidad. Consideramos que si no se atiende este aspecto puede tener como resultado que el programador abandone su uso. También es importante que se proporcione documentación en otros idiomas además del inglés y chino, ya que esto

fomentaría su rápida adopción en la comunidad académica y científica.

Finalmente, en relación con el sector primario, a fin de que los agricultores de palma datilera, pudieran comprender claramente el alcance que tiene actualmente la inteligencia artificial y como está facilitando las tareas en el mundo moderno, el principal reto para este grupo de investigación, fue ejemplificar de la forma más sencilla y clara posible, la terminología y cultura computacional. Por otra parte, los productores de dátil, mostraron una total disposición y ofrecieron todas las facilidades para el desarrollo de nuestro proyecto.

5. Contribución del proyecto

La principal contribución del proyecto fue el desarrollo de aplicaciones de inteligencia artificial centradas en las necesidades y problemáticas de los pequeños productores de dátil en México. Estas tienen el objetivo de ayudar a los productores a reducir costos, optimizar procesos, incrementar productividad y sustentabilidad.

Otras contribuciones fueron establecer lazos de colaboración entre las universidades y los sectores productivos del país, desarrollar capacidades digitales y generar capital humano especializado en tecnologías emergentes e inteligencia artificial.

El desarrollo de estas actividades, se considera que tendrán un fuerte impacto social a mediano y largo plazo, beneficiando en mayor medida al sector primario del país.

6. Conclusiones

Es importante que las empresas de tecnología de punta generen lazos de colaboración que tengan por objetivo el apoyar e incentivar a los académicos, investigadores y sociedad en general, para desarrollar proyectos de innovación tecnológica con énfasis en inteligencia artificial (IA) y tecnologías emergentes. En donde se propongan proyectos que ofrezcan soluciones al sector primario, problemas sociales o ambientales y que contribuyan al desarrollo económico del país.

El participar en la sinergia ofrecida por la convocatoria Alianza-Huawei UNAM³ para el equipo ha sido enriquecedor. Debido a que nos ha permitido realizar un prototipo funcional del proyecto.

¹<https://www.docker.com/>

²<https://e.huawei.com/mx/products/cloud-computing-dc/atlas/mindspore>

³<https://alianza.unam.mx/>

BIBLIOGRAFÍA

- [1] A. Manickavasagan, N. K. Al-Mezeini y H. N. Al-Shekaili, "RGB color imaging technique for grading of dates", *Scientia Horticulturae*, vol. 175, pp. 87–94, agosto de 2014. Accedido el 9 de septiembre de 2022. [En línea]. Disponible: <https://doi.org/10.1016/j.scienta.2014.06.003>
- [2] H. Altaheri, M. Alsulaiman, G. Muhammad, S. U. Amin, M. Bencherif y M. Mekhtiche, "Date fruit dataset for intelligent harvesting", *Data in Brief*, vol. 26, p. 104514, octubre de 2019. Accedido el 9 de septiembre de 2022. [En línea]. Disponible: <https://doi.org/10.1016/j.dib.2019.104514>
- [3] K. Albarrak, Y. Gulzar, Y. Hamid, A. Mehmood y A. B. Soomro, "A Deep Learning-Based Model for Date Fruit Classification", *Sustainability*, vol. 14, n.º 10, p. 6339, mayo de 2022. Accedido el 10 de septiembre de 2022. [En línea]. Disponible: <https://doi.org/10.3390/su14106339>
- [4] M. Schmeisser et al., "Parallel, distributed and GPU computing technologies in single-particle electron microscopy", *Acta Crystallographica Section D Biological Crystallography*, vol. 65, n.º 7, pp. 659–671, junio de 2009. Accedido el 10 de septiembre de 2022. [En línea]. Disponible: <https://doi.org/10.1107/s0907444909011433>
- [5] H. Liao, J. Tu, J. Xia y X. Zhou, "DaVinci: A Scalable Architecture for Neural Network Computing", en 2019 IEEE Hot Chips 31 Symposium (HCS), Cupertino, CA, USA, 18–20 de agosto de 2019. IEEE, 2019. Accedido el 10 de septiembre de 2022. [En línea]. Disponible: <https://doi.org/10.1109/hotchips.2019.8875654>

Fecha de recepción: julio, 2023

Fecha de publicación: noviembre de 2023

Cómo se cita:

J. P. García, R. Salomón, D. B. Pérez, H. Zatarain, J. E. Soto, E. R. García, "Inteligencia Artificial al Servicio de los Pequeños Productores de Dátil en México," *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 8, noviembre, 2023. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].