

TIES Revista de Tecnologías de Información en Educación Superior

Publicación Semestral • Octubre de 2021 • ISSN en trámite

Editorial

Fernando Arámbula Cosío



De redes neuronales recurrentes a modelos de lenguaje:
la evolución del PLN en la generación de textos

David Emmanuel Maqueda Bojorquez



Clasificación de poblaciones nativas de frijol utilizando visión artificial

José Luis Morales-Reyes, Héctor Gabriel Acosta-Mesa, Elia Nora Aquino-Bolaños,
Socorro Herrera-Meza, Nicandro Cruz-Ramírez y José Luis Chávez-Servia



Las anomalías: ¿qué son?, ¿dónde surgen?, ¿cómo detectarlas?

Nidiyare Hevia-Montiel, Sergio Mota y Antonio Neme



Aprendizaje computacional para análisis de imágenes de ultrasonido médico

F. Arámbula Cosío, F. Torres Robles, G.A.R. Velásquez-Rodríguez, E. Galicia Gómez,
B. Escalante-Ramírez, J. Olveres, Jorge L. Pérez, V. Medina Bañuelos, L. Camargo Marín y
M. Guzmán Huerta



Aprendizaje computacional aplicado a la detección de huesos, en cirugía
ortopédica asistida por computadora

Zian Fanti, Alfonso Gastelum Strozzi, Eric Hazan Lasri y F. Arámbula Cosío



TIES, REVISTA DE TECNOLOGÍA E INNOVACIÓN EN EDUCACIÓN SUPERIOR (www.ties.unam.mx) 2021, Año 2, No. 4, octubre 2021, es una publicación semestral editada por la Universidad Nacional Autónoma de México (UNAM), Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, Ciudad de México, a través de la Dirección General de Cómputo y de Tecnologías de Información y Comunicación, (DGTIC), Circuito Exterior s/n, Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, Ciudad de México, Teléfono: (55) 56228166, <http://www.ties.unam.mx>, revista.ties@unam.mx. Editor responsable: Lic. Lizbeth Luna González. Número de reserva de Derechos de Autor otorgado por INDAUTOR: 04-2019-011816190900-203 ISSN: en trámite, ambos otorgados por el Instituto Nacional del Derecho de Autor. Responsable de la última actualización de este número, Dirección General de Cómputo y de Tecnologías de Información y Comunicación, (DGTIC). Circuito Exterior s/n, Ciudad Universitaria, Alcaldía Coyoacán, C.P. 04510, Ciudad de México, fecha de la última modificación, octubre de 2021.

El contenido de los artículos es responsabilidad de los autores y no refleja el punto de vista de los árbitros, del Editor o de la UNAM. Se autoriza la reproducción total o parcial de los textos aquí publicados siempre y cuando se cite la fuente completa y la dirección electrónica de la publicación. La revista se ha desarrollado sin fines de lucro, con finalidades de disseminación del conocimiento, bajo licencia Creative Commons Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0). Hecho en México, 2021.



DIRECTORIO

Universidad Nacional Autónoma de México

Dr. Enrique Luis Graue Wiechers
Rector

Dr. Leonardo Lomelí Vanegas
Secretario General

**Dirección General de Cómputo y Tecnologías
de la Información y Comunicación**

Dr. Héctor Benítez Pérez
Director de la DGTIC

José Fabián Romo Zamudio
**Director de Sistemas y Servicios
Institucionales**

M.A.O. Miguel Ángel Mejía Argueta
Responsable de Acervos Digitales

Lic. Lizbeth Luna González
Directora Editorial de la Revista

Directorio histórico

Dr. Felipe Bracho Carpizo
Director Fundador de la revista (2018-2020)

Dr. Luis Alberto Gutiérrez Díaz de León
(Q.E.P.D.)
Director General de la revista (2018-2020)

CRÉDITOS

Dr. Héctor Benítez Pérez
Director General de la Revista

Lic. Lizbeth Luna González
Directora Editorial de la Revista

Liliana Minerva Mendoza Castillo
Asistente Editorial de la Revista

José Fabián Romo Zamudio
Lic. Lizbeth Luna González
Liliana Minerva Mendoza Castillo
Diseño de Contenidos

Lic. Lizbeth Luna González
Arquitectura de la información

Mtro. Rodolfo Cano Ramírez
**Diseño de interfaz gráfica y
Formación PDF**

Lic. Adrián Estrada Corona
Liliana Minerva Mendoza Castillo
Corrección de Estilo

Mtro. Rodolfo Cano Ramírez
Estructura HTML

Ing. Carlos Alberto Román Zamitiz
D.I. Óscar Isaías Del Río Martínez
Abigail Sánchez Gálvez
Lic. Lizbeth Luna González
Liliana Minerva Mendoza Castillo
MAO. Miguel Ángel Mejía Argueta
Formación HTML

Lic. Lizbeth Luna González
Ing. Carlos Alberto Román Zamitiz
Liliana Minerva Mendoza Castillo
Abigail Sánchez Gálvez
M.A.O. Miguel Ángel Mejía Argueta
Administrador del OJS

Ing. Carlos Alberto Román Zamitiz
Abigail Sánchez Gálvez
Administrador del servidor

COMITÉ EDITORIAL

Dr. Héctor Benítez Pérez, Presidente, Universidad
Nacional Autónoma de México. MÉXICO

María de Lourdes Velázquez Pastrana, Encargada del
Despacho, Universidad Nacional Autónoma de México.
MÉXICO

Alonso Castro Mattei, Universidad de Costa Rica.
COSTA RICA

Ernesto Chinkes, Universidad de Buenos Aires.
ARGENTINA

Carmen Humberta de Jesús Díaz Novelo, Universidad
Autónoma de Yucatán. MÉXICO

Salma Jalife Villalón, Corporación Universitaria para el
Desarrollo de Internet. MÉXICO

Lizbeth Luna González, Universidad Nacional
Autónoma de México. MÉXICO

José Luis Ponce López, Asociación Nacional de
Universidades e Instituciones de Educación Superior.
MÉXICO

Raúl Rivera Rodríguez, Centro de Investigación
Científica y Educación Superior de Ensenada. MÉXICO

José Fabián Romo Zamudio, Universidad Nacional
Autónoma de México. MÉXICO

Luis Alberto Gutiérrez Díaz de León (Q.E.P.D.),
Secretario, Universidad de Guadalajara. MÉXICO

Felipe Bracho Carpizo, Presidente, Universidad
Nacional Autónoma de México. MÉXICO (2018-2020)

ÍNDICE

| | |
|--|----|
| Editorial | 5 |
| Fernando Arámbula Cosío | |
| De redes neuronales recurrentes a modelos de lenguaje: la evolución del PLN en la generación de textos | 6 |
| David Emmanuel Maqueda Bojorquez | |
| Clasificación de poblaciones nativas de frijol utilizando visión artificial | 16 |
| José Luis Morales-Reyes, Héctor Gabriel Acosta-Mesa, Elia Nora Aquino-Bolaños, Socorro Herrera-Meza, Nicandro Cruz-Ramírez y José Luis Chávez-Servia | |
| Las anomalías: ¿qué son?, ¿dónde surgen?, ¿cómo detectarlas? | 27 |
| Nidiyare Hevia-Montiel, Sergio Mota y Antonio Neme | |
| Aprendizaje computacional para análisis de imágenes de ultrasonido médico | 38 |
| F. Arámbula Cosío, F. Torres Robles, G.A.R. Velásquez-Rodríguez, E. Galicia Gómez, B. Escalante-Ramirez, J. Olveres, Jorge L. Pérez, V. Medina Bañuelos, L. Camargo Marín y M. Guzmán Huerta | |
| Aprendizaje computacional aplicado a la detección de huesos, en cirugía ortopédica asistida por computadora | 52 |
| Zian Fanti, Alfonso Gastelum-Strozzi, Eric Hazan-Lasri y F. Arámbula Cosío | |

EDITORIAL

En años recientes el concepto de Inteligencia Artificial (IA) se ha difundido ampliamente en todo tipo de medios de comunicación. Actualmente, un gran número de aparatos y programas de computadora (software) incluyen algún tipo de función automática que replica alguna función humana: reconocimiento y generación de lenguaje natural; planeación de rutas en un mapa; reconocimiento de sonrisas en las fotos. Dentro de los proyectos de aplicación de técnicas de IA tenemos: automóviles autónomos y robots humanoides capaces de caminar, de interactuar con personas y de realizar tareas dentro de ambientes no estructurados (p. ej. dentro de hospitales). En muchos casos se utilizan redes neuronales profundas (*deep neural networks*) para realizar funciones humanas. Por ejemplo, las redes neuronales convolucionales se utilizan en secuencias de videos para detectar automáticamente objetos, animales y personas; también se pueden usar en robots, vehículos autónomos, cámaras de seguridad y análisis automático de imágenes médicas.

Las redes neuronales profundas han demostrado gran capacidad para la solución de problemas complejos (lenguaje, visión, planeación), y su amplia aplicación podría sugerir que la IA se constituye de diferentes arquitecturas de redes neuronales profundas. Sin embargo, en 1956 se le consideró como un área de investigación multidisciplinaria integrada por: ciencias cognitivas, electrónica, computación y matemáticas. Desde entonces, se ha desarrollado la teoría y las implementaciones prácticas para replicar algunas funciones humanas en computadora, dentro de áreas de estudio como: reconocimiento

de patrones, aprendizaje computacional o aprendizaje automático, visión artificial, procesamiento de lenguaje natural y, recientemente, ciencia de datos.

TIES, Revista de Tecnología e Innovación en Educación Superior, incluye en este número, cinco trabajos de IA que presentan un panorama de los diferentes conocimientos matemáticos y computacionales que se utilizan para la implementación de funciones humanas en computadora. El primer artículo, “Las anomalías: ¿qué son?, ¿dónde surgen?, ¿cómo detectarlas?”, nos presenta un análisis de las anomalías que pueden ocurrir en los datos de entrenamiento y cómo el proceso adecuado puede contribuir a desarrollar mejores algoritmos de aprendizaje. El segundo artículo, “De redes neuronales recurrentes a modelos de lenguaje: la evolución del PLN en la generación de textos”, describe con detalle la evolución del procesamiento de lenguaje natural y la generación de texto basados en redes neuronales desde los años noventa. El trabajo: “Clasificación de poblaciones nativas de frijol utilizando visión artificial”, aborda la aplicación de técnicas de visión artificial en agricultura. Los dos artículos finales, “Aprendizaje computacional para análisis de imágenes de ultrasonido médico” y “Aprendizaje computacional aplicado a la detección de huesos, en cirugía ortopédica asistida por computadora”, aplican las técnicas de reconocimiento de patrones y aprendizaje automático en el análisis de imágenes médicas.

Dr. Fernando Arámbula Cosío
Editor invitado

DE REDES NEURONALES RECURRENTE A MODELOS DE LENGUAJE: LA EVOLUCIÓN DEL PLN EN LA GENERACIÓN DE TEXTOS

David Emmanuel Maqueda Bojorquez

<http://www.ties.unam.mx/>

Fecha de recepción: julio 03, 2021 • Fecha de publicación: octubre de 2021

Octubre 2021 | número de revista 4 • ISSN en trámite



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)

DE REDES NEURONALES RECURRENTE A MODELOS DE LENGUAJE: LA EVOLUCIÓN DEL PLN EN LA GENERACIÓN DE TEXTOS

Resumen

En el presente texto haremos un breve repaso por la evolución del procesamiento de lenguaje natural y la generación de texto artificial. Desde la década de 1990, vieron la luz las primeras ideas sobre generación artificial de textos. En una primera instancia, las técnicas existentes resultaron insuficientes para resolver tareas, tales como la propia generación de textos, la traducción automática o la búsqueda de información.

Sin embargo, al pasar del tiempo se fueron generando nuevas ideas, con base en la comprensión propia de la generación de textos desde la naturaleza humana, desde los primeros modelos de redes neuronales recurrentes hasta los modelos de lenguaje más actuales, desde la consecución de palabras más básicas hasta lograr la generación de obras literarias enteras de manera artificial. De esta forma visualizaremos en esencia cómo es que se dio esta evolución en el campo y cuál fue la lógica que siguió.

Palabras clave:

Procesamiento de Lenguaje Natural, Modelos de Lenguaje, Redes Neuronales Recurrentes, LSTM, Inteligencia Artificial

FROM RECURRING NEURAL NETWORKS TO LANGUAGE MODELS: THE EVOLUTION OF NLP FOR TEXT GENERATION

Abstract

In this text we will briefly review the evolution of natural language processing and the generation of artificial text. Since the 1990s, the first ideas about artificial text generation saw the light; In the first instance, the existing techniques were insufficient to solve tasks such as the generation of texts, automatic translation or the search of information.

However, as time went by, new ideas were generated based on the understanding of the generation of texts from human nature, from the first models of recurrent neural networks to the most current language models, from the achievement of words more basic until the generation of entire literary works artificially. In this way we will visualize in essence how this evolution occurred in the field and what was the logic that said evolution followed.

Keywords:

Natural Language Processing, Language Models, Recurrent Neural Networks, LSTM, Artificial Intelligence.

DE REDES NEURONALES RECURRENTE A MODELOS DE LENGUAJE: LA EVOLUCIÓN DEL PLN EN LA GENERACIÓN DE TEXTOS

Introducción

Procesamiento de Lenguaje Natural (PLN) es la parte de la Inteligencia Artificial (IA) y lingüística que interactúa con: el lenguaje humano, hablado o escrito y la computación [1]. Para llevar a cabo este proceso es distintivo el uso del lenguaje en general, se requiere tener estudios de fonética, fonología, morfología, semántica y sintaxis.

Además, se necesita acercarse al significado léxico, a la semántica composicional, para entender el concepto de cada palabra. Para interpretar las intenciones del lenguaje, se debe estudiar pragmática y diálogo. Finalmente, para deducir y asociar información dentro del lenguaje, se necesita comprensión del discurso. [2]

Este conjunto de conocimiento lingüístico se intenta brindar a un sistema, mediante un modelo matemático/computacional, que depende claramente del tipo de procedimiento a desarrollar. Mientras un sistema conversacional (e.g. Siri, Cortana o Alexa), con clara tendencia al reconocimiento y síntesis de voz, necesitaría una amplia cognición de fonética y fonología, ya que interpretaría y generaría voz. Otro tipo de generadores de texto necesitaría mayor noción de la sintaxis, la ortografía, el contexto léxico y la estructura de un lenguaje, idioma o dialecto, [3] siendo dependientes al igual que su fin, debido a las discre-

pancias entre el estilo de escritura que puede haber entre estilos de textos independientes de la lengua (e.g. escritura poética, científica o coloquial).

En este punto cabe destacar la prueba de Turing, que él mismo llamó ‘*The Imitation Game*’ o el juego de imitación. [4] En esencia, esta prueba dice que es suficiente usar únicamente lenguaje para probar si una máquina se puede considerar pensante, todo esto logrando confundir a una persona mediante el origen de un cierto mensaje. Si el interlocutor considera que un mensaje, o respuesta a una pregunta, no es generado artificialmente, entonces el agente artificial puede considerarse inteligente. Es llamado el juego de imitación, ya que el agente artificial imita al ser humano. Para completar el experimento se contrastan las respuestas de un humano y un agente artificial con un tercero, el que desconoce la fuente de los mensajes que recibe.

Para muchos, la habilidad de procesar lenguaje como los humanos es un signo claro de máquinas inteligentes. Así que la tendencia en el campo ha ido en dirección a mejorar la forma en que se generan textos y lenguaje en general, viéndose mejorada la calidad de estos. Los avances recientes en sistemas conversacionales muestran a las personas de cierta forma. Comienzan a tratar a las máquinas como personas, incluyéndolas directamente en su vida cotidiana. [5]

Desarrollo

Un breve repaso sobre el funcionamiento de redes neuronales

Actualmente, la mayoría de los desarrollos en IA están fundamentados en redes neuronales, modelos matemáticos / computacionales que provienen de los estudios relacionados con el funcionamiento del cerebro. En 1943 McCulloch y Pitts aportaron una forma de describir las funciones cerebrales en términos abstractos, para mostrar que los elementos simples, conectados en una red neuronal, pueden tener un inmenso poder computacional. [6] Como lo indica su nombre, el concepto es crear una red de neuronas. Las neuronas por sí solas son unidades de cálculo, por las cuales fluye información (imágenes, textos, etcétera, codificados en vectores, los cuales son objetos matemáticos fácilmente manejables). Al final de esta tendremos una salida resultante, dependiente de los datos que fluyen por la red.

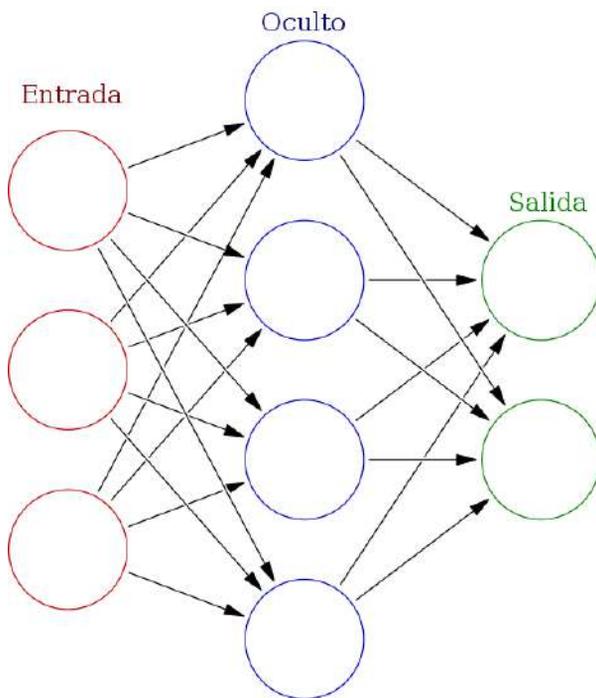


Figura 1. Estructura general de una red neuronal. Disponible en: https://commons.wikimedia.org/wiki/File:Colored_neural_network_es.svg. Consultado en junio 02, 2021.

En este punto, cabe destacar el concepto de capa neuronal. Como previamente se describió de manera breve, una neurona es una unidad de cálculo que también podría definirse como el elemento básico dentro de una red neuronal, mientras que una capa neuronal es en sí un conjunto de neuronas que tienen características homogéneas entre ellas.

Dentro de una red neuronal, la información fluye de dos formas y en ambas tiene utilidades distintas. Cuando alimentamos la red (*forward*), la información fluye de la generalidad a la particularidad (de la entrada de la red a la salida de la misma), es decir, la red neural, así como cada una de las neuronas, se va contextualizando matemáticamente con respecto a los datos que fluyen por cada una de las unidades de cálculo (neuronas). [7]

Cuando la información fluye a la salida, el siguiente tipo de flujo de información es el *backward*, que actúa sobre la salida y genera una retroalimentación a toda la red neuronal, relacionada con cómo mejorar la calidad de la salida, modificando los datos con que se calcula el valor de cada neurona. Esta modificación se hace en base a una resultante, entre la diferencia del valor esperado y lo que nos es ofrecido por la red neural. A esto se le denomina *error o pérdida* (hablando de aprendizaje supervisado). [8]

En la figura 1 se muestra la estructura general de una red neural. Se presentan gráficamente las neuronas de cada una de las capas de la red. Las neuronas de la capa de entrada son las que reciben la información directamente, de los datos con que se alimenta a la red. Esto significa que desde ahí inicia el *forward* de la red, mientras que a partir de la capa de salida se calcula un error, en base a la salida de la red, e inicia el *backward*. Cabe destacar que, típicamente, una capa oculta se define como toda aquella que se encuentre entre la capa de entrada y la de salida. Puede haber una o varias capas ocultas en una arquitectura neuronal. En el caso del diagrama, se visualiza entonces una capa oculta. Las neuronas, una a una, se conectan con fuerzas variables de conexión (o ponderaciones), comúnmente llamadas pesos.

A grandes rasgos, mediante el proceso antes descrito se entrena una red neuronal, es decir, la forma a través de la cual se va refinando un agente inteligente para generar mejores resultados. Cada vez que este proceso se repite con cada uno de los datos que están a disposición, se determina que se ha cumplido una *época* de entrenamiento. [9]

El concepto de memoria: Redes Neuronales Recurrentes

La idea de la generación de textos no es nueva en lo absoluto, si bien es cierto que los primeros esfuerzos se dieron en la década de los cuarenta con los n-gramas, el proyecto se remonta a los experimentos realizados por Claude Shannon, [10] en que la idea era: dada una secuencia de letras, ¿cuál es la siguiente letra más probable? Como tal, siendo un modelo probabilístico, los n-gramas permiten predicciones estadísticas del próximo elemento de una cierta secuencia de elementos, donde, debido a limitaciones computacionales y a la normalmente naturaleza abierta de los problemas (suele haber infinitos elementos posibles), se asume que cada elemento solo depende de los últimos n elementos de la secuencia. [11]

Alrededor de la década de los noventa, considerando las limitaciones de los n-gramas y, además, con los resultados prometedores de las redes neuronales tradicionales,

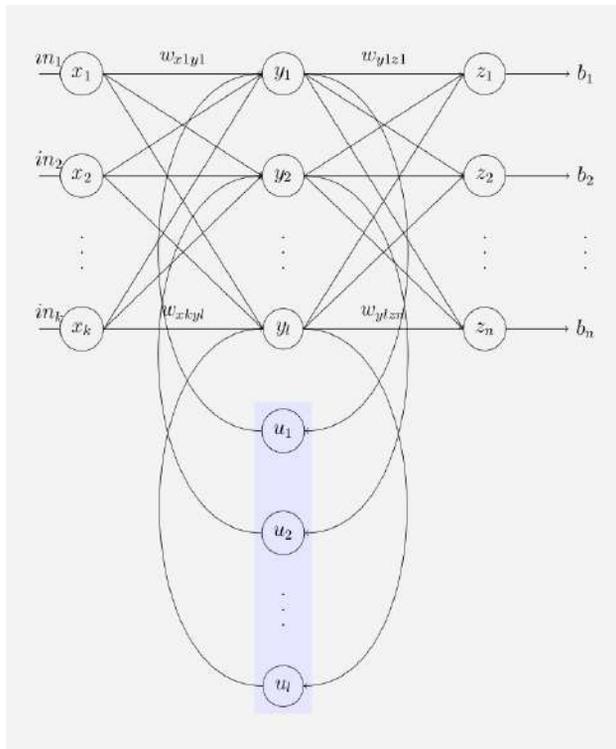


Figura 2. Estructura de la red neuronal recurrente simple de Elman. Disponible en: https://en.wikipedia.org/wiki/Recurrent_neural_network#/media/File:Elman_srnn.png Consultado en junio 02, 2021.

feed-forward, surgen las redes neuronales recurrentes o *Recurrent Neural Networks (RNN's)*. El principio básico de este tipo de redes se basa en la utilización de bucles dentro del flujo de información. Recordemos que en las redes neuronales convencionales solo existe el *backward* y el *forward*, para describir la naturaleza del flujo; sin embargo, como puede verse en la figura 2, en una *RNN* se incluyen bucles en el flujo, los cuales están representados por las líneas curvadas intermedias en el diagrama. Dichos bucles de información son interpretados como remanentes de información, que realimentan a la red junto con información nueva, que viene desde la entrada. Esto permite almacenar información dentro de la red.

En resumen, las *RNN* utilizan su razonamiento de experiencias anteriores para informar los próximos eventos. Un ejemplo común de *RNN* es la traducción automática. Por ejemplo, una red neuronal puede tomar una oración de entrada en español y traducirla a una oración en inglés. La red determina la probabilidad de cada palabra en la oración de salida, basándose en la palabra en sí y en la secuencia de salida anterior. [12]

Sin embargo, uno de los principales problemas de las *RNN* fue su limitado rango de información contextual, es decir, estos sistemas pueden considerar la información contextual de una situación previa, más no de la generalidad. En un contexto de traducción, esto fue problemático, debido a que en la mayoría de las situaciones la correcta traducción no depende exclusivamente de la palabra antecedente, sino también de un contexto mucho más generalizado en la oración o párrafo a traducir.

En conclusión, esto, aunado a problemas relacionados con el entrenamiento de estas redes, creó la situación idónea para desarrollar nuevas ideas. Así es como en 1997 surgen las redes *Long Short-Term Memory (LSTM)*. [13]

Introducción a los tipos de memoria: Long Short-Term Memory

El concepto de una *LSTM* sigue siendo una red recurrente, con los bucles de flujo de información, pero aplica transformaciones sofisticadas a las entradas. Las entradas de cada neurona se manipulan mediante maniobras complejas, lo que produce dos salidas, que se pueden considerar como la *memoria a largo plazo* y la *memoria a corto plazo*, dos bucles de flujo de información distintos.

Esta adición de un flujo de información a largo plazo amplía drásticamente el tamaño de la atención de la red.

Puede acceder a estados neurales anteriores y así obtener un mayor contexto que una *RNN*, sin embargo, la red *LSTM* tiene sus desventajas. Sigue siendo una red recurrente. Si bien la adición de un canal de memoria a largo plazo ayuda, existe un límite en la cantidad que puede contener. [14]

En la figura 3, se muestra la visualización de una celda de la red *LSTM*, compuesta por una compuerta de entrada, una puerta de salida y una compuerta de olvido. La celda recuerda valores en intervalos de tiempo arbitrarios y las tres puertas regulan el flujo de información dentro y fuera de la celda. De esta forma, se pueden procesar datos secuencialmente y mantener su estado oculto a lo largo del tiempo. [15]

La *LSTM* funcionó bien durante un tiempo. Pudo generar caracteres razonablemente bien en textos más cortos y no se dejó intimidar por muchos de los problemas que plagaron el desarrollo temprano del procesamiento del lenguaje natural, en particular, una mayor profundidad global y comprensión, no solo de palabras individuales, sino también de su significado colectivo.

Las *LSTM* son limitadas cuando se presentan problemas modernos más exigentes, como la traducción automática en varios idiomas o la generación de texto completamente indistinguible del texto escrito por humanos. Cada vez más, se utiliza una arquitectura más nueva para abordar tareas más desafiantes: *Transformers*.

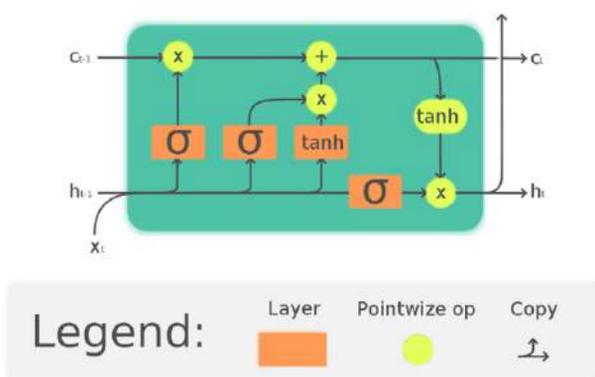


Figura 3. Celda de LSTM para procesar secuencialmente la información. Su información se puede mantener oculta en el tiempo. Disponible en: https://en.wikipedia.org/wiki/Recurrent_neural_network#/media/File:Elman_s_rnn.png Consultado en junio 02, 2021.

Lo único que necesitas es atención: transformers

Los *transformers* o transformadores son un modelo introducido en 2017, [16] que se denomina arquitectura secuencia a secuencia. Secuencia a secuencia (o Seq2Seq) es una red neuronal que transforma una secuencia determinada de elementos (como la secuencia de palabras en una oración) en otra secuencia. Por ejemplo, palabras de un idioma a una secuencia de palabras diferentes en otro idioma.

Transformers hace uso de un concepto sumamente importante: la 'atención'. El mecanismo de atención mira una secuencia de entrada y decide en cada paso qué otras partes de la secuencia son importantes. Por ejemplo, al leer este texto, siempre se concentra en la palabra que lee, pero, al mismo tiempo, su mente aún guarda las palabras clave importantes del texto en la memoria, para proporcionar contexto.

Las redes recurrentes eran, hasta ahora, una de las mejores formas de capturar las dependencias oportunas en secuencias. Sin embargo, el equipo que presentó el artículo demostró que una arquitectura, con solo mecanismos de atención, puede mejorar los resultados en la tarea de traducción y otras tareas. Los *transformers* no requieren que los datos secuenciales se procesen en orden. Más bien, la operación de atención identifica el contexto para cualquier posición en la secuencia de entrada.

Al usar 'atención' se pondera la importancia de cada elemento de la secuencia de entrada, es decir, identifica el contexto que confiere significado a una palabra en la oración. Al dejar atrás el paradigma principal de las *RNN*, sobre los bucles de flujo de información y el concepto de memoria a corto y largo plazo, los *transformers* se han convertido rápidamente en el modelo de elección para los problemas de *PLN*, reemplazando los modelos de *RNN* más antiguos.

Modelos de Lenguaje, la implementación de Transformers

Un modelo de lenguaje se ha vuelto recientemente en un eje central del *PLN*, el cual analiza el patrón del lenguaje humano para la predicción de palabras. [17] Las aplicaciones basadas en *PLN* utilizan modelos de lenguaje para una variedad de tareas, como conversión de audio a texto, reconocimiento de voz, análisis de sentimientos, resumen, corrección ortográfica, etcétera.

Los modelos de lenguaje determinan la probabilidad de la siguiente palabra, analizando el texto en los datos. Los modelos se preparan para la predicción de palabras, aprendiendo las características de un idioma. Con este aprendizaje, el modelo se prepara para comprender frases y predecir las siguientes palabras en frases. La cantidad de datos de texto que se analizarán y las matemáticas aplicadas para el análisis, marcan la diferencia en el enfoque seguido para crear y entrenar un modelo de lenguaje.

Además de las ventajas que supone la implementación de los *Transformers* antes mencionados, también permite mucha más paralelización que los *RNN* convencionales. Esto, por lo tanto, reduce los tiempos de entrenamiento, [18] lo que ha permitido el desarrollo de modelos de lenguaje, como *BERT*, *GTP*, *GTP-2* y el recientemente introducido *GTP-3*, entrenados con conjuntos de datos de lenguaje general de gran tamaño, como *Wikipedia Corpus* (1.9 billones de palabras) [19] y *Common Crawl* (220 TB de información), [20] que pueden ser adaptados a tareas lingüísticas específicas y cuyos beneficios son mucho mayores de lo que era posible obtener antes de la llegada de *Transformers* en 2017.

Un modelo de lenguaje es, a grandes rasgos, una distribución de probabilidad de secuencias de palabras. ¿Por qué esto es necesario y a qué nos referimos?

Bien, supongamos que tenemos una serie de palabras: *mesa, sobre, plato, la, el, está*. Con estas palabras tendremos que generar una oración. Para un humano que tenga conocimiento del idioma español, resultará naturalmente sencillo generar la oración: “*El plato está sobre la mesa*”. Sin embargo, para una máquina la historia es distinta. Un modelo de lenguaje, usando la distribución de probabilidad, tendrá que definir cuál sucesión de palabras tiene la mejor probabilidad entre todas las posibilidades, a lo que un modelo con un entrenamiento correctamente realizado llegaría a una conclusión igual o similar a la de un humano.

Esto quiere decir que la sucesión de palabras “*El plato está sobre la mesa*”, tiene una mayor distribución de probabilidad que la sucesión “*La mesa está sobre el plato*” o “*Sobre el plato, está el mesa*”. Ahora bien, aquí podría surgir la pregunta: ¿cómo se logra esto?

Como dijimos antes, estos modelos, al ser entrenados con conjuntos de datos tan grandes, tienen gran contexto del manejo del lenguaje. Mientras más grandes sean estos conjuntos de datos, mayor será su conocimiento del lenguaje y el manejo de este (con el debido tiempo de entrenamiento). Esto significa que, al ya ser *refinado* el

modelo con una gran cantidad de datos y recordando el funcionamiento del método de ‘atención’ propuesto por *Transformers*, se abstrae la información contextual de cada palabra en una oración. Esto quiere decir que el modelo de lenguaje tendrá conocimiento de la importancia de cada elemento, de la secuencia de datos, con la que haya sido entrenado, [21] es decir, de cada una de las oraciones, de cada párrafo, de cada documento dentro de toda la información con la que haya sido entrenado el modelo de lenguaje. Es por ello por lo que entre mayor sea el conjunto de datos con el cual se entrena, mayores serán los fenómenos lingüísticos de los que el modelo tendrá contexto y, por ello, el mismo será más robusto.

Por este motivo, el modelo puede deducir cuál secuencia de palabras es la de mayor distribución de probabilidad. Resulta más probable que la secuencia “*El plato*” o “*La mesa*”, sean lingüísticamente correctas, que “*La plato*” o “*El mesa*”, dado que todos los eventos lingüísticos, de los cuales el modelo obtuvo contexto, eran del primer tipo, dado que la forma correcta del uso del lenguaje es ésta, mientras la otra es improbable, dado que no es la forma correcta de uso de lenguaje, por lo que los textos, de los que el modelo obtuvo contexto, no contenían estos tipos de secuencias.

En el caso de generación de textos, al introducir una secuencia inicial (*prompt*), ésta será la base que calcule una distribución de probabilidad, es decir, la predicción de la palabra consecuente estará directamente relacionada no solamente con la palabra antecedente (como habría sido en las *RNN* convencionales), sino también con toda la información contextual que brinden las demás palabras, que incluyan la secuencia inicial. Cuando esta nueva palabra sea generada, será considerada nuevamente para la generación de otra nueva palabra, así como del nuevo contexto que le haya dado a la secuencia completa y así sucesivamente.

Cada nueva palabra generada depende del contexto de la secuencia base y también, a su vez, del contexto que tenga el modelo del lenguaje, dado su entrenamiento. Volviendo al supuesto anterior, al tener la frase “*El plato*”, aunque puedan existir una infinidad de palabras consecuentes lingüísticamente correctas, los datos y a su vez los fenómenos lingüísticos con los cuales el modelo fue entrenado, decidirán cuál es la palabra consecuente con mayor probabilidad. Supongamos que dicha palabra es *está*, siendo la nueva secuencia base “*El plato está*”. En este punto la información contextual de la secuencia base se modificó: *el plato debe estar en algún lugar*. Así que se delimita la palabra a generar, con alguna que tenga sentido contex-

tual, más allá de una correcta sintaxis. El modelo incluye esa información en la generación de una nueva palabra y, dada la forma como se constituye el lenguaje (en este caso el español), analizando tanto el contexto global del entrenamiento (conocimiento general del lenguaje del modelo), como el de la secuencia base, se generará una palabra. Supondremos que las siguientes dos palabras generadas serán “en la”, siendo la nueva secuencia base “*El plato está en la*”. De nueva cuenta el contexto de la oración cambia, ya que *el plato debe estar en un objeto cuyo prefijo sea femenino*, delimitando así la generación de la nueva palabra. Pueden existir palabras compatibles, como “*nevera*”, “*alacena*”, etcétera. Sin embargo, aunque todas son lingüísticamente correctas, supongamos que el modelo decida que la palabra más probable es: “*mesa*”. Esto significa que es más común y, a su vez, más probable que la palabra más adecuada en el idioma español, o al menos en el conjunto de entrenamiento del modelo, sea esa y con esto terminamos con el ejercicio supuesto.

Conclusiones

Como puede notarse con estos modelos, el razonamiento y la generación de textos es similar a la forma como un humano generaría algún texto. Incluso, al estar escribiendo este texto, se está usando el mismo principio: se generó una idea, se inició una frase y con base en esa frase se construyó un texto contextualmente coherente y sintácticamente correcto, que represente la idea que se quiso expresar al momento de escribir. La diferencia radica en la generación inicial de la idea. Un humano, como cualquiera de nosotros, puede generarla intrínsecamente. Un modelo como el expuesto previamente, es dependiente de una secuencia de palabras inicial, que debe brindársele, aunque el texto de esa idea inicial y el rumbo de este es independiente de una idea explícitamente humana.

Mientras un modelo de lenguaje conocería la sintaxis del idioma, la coherencia contextual y la precisión léxica, con base en un conjunto de datos de entrenamiento y el uso del lenguaje, un humano lo conocería con base en años de interacción humana y un adiestramiento riguroso.

Recursos Adicionales

Demo con transformers: [En línea]. Disponible en <https://transformer.huggingface.co/> [Consultado en junio 02, 2021].

BIBLIOGRAFÍA

- [1] A. Gelbukh, "Procesamiento del lenguaje natural: estado de la investigación," en *Memoria del I Simposio Internacional sobre Organización del Conocimiento: Bibliotecología y Terminología*, C. Naumis Peña, 2009, pp. 337-357.
- [2] G. Guida y G. Mauri, "Evaluación de los sistemas de procesamiento del lenguaje natural: problemas y enfoques," en *Proceedings of the IEEE*, vol. 74, no.7, pp. 1026–1035, 1986.
- [3] T. Klüwer, "From chatbots to dialog systems," en *Conversational agents and natural language interaction: Techniques and Effective Practices*, 2011, pp.1-22.
- [4] A. M. Turing, "Computing Machinery and Intelligence," en *Mind*, 1950, pp. 433-460.
- [5] D. Jurafsky & J.H Martin, *Speech and language processing*. Pearson International Edition, 2009 ISBN 978-0-13-504196-3, Chapter 24.
- [6] W.S. McCulloch y W.H Pitts, "A logical Calculus of ideas immanent in nervous activity," en *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133, 1943.
- [7] O. Varun Kumar, A. Ajith y S. Václav, "Metaheuristic design of feedforward neural networks: A review of two decades of research," en *Engineering Applications of Artificial Intelligence*, vol. 60, pp. 97-116, 2017.
- [8] S. Hochreiter y J. Schmidhuber, "Long short-term memory," en *Neural Computation*, vol. 9 no.8, pp. 1735–1780, 1997.
- [9] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. New York: Macmillan College Publishing, 1998.
- [10] C.E. Shannon, "A Mathematical Theory of Communication," en *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.
- [11] C. D. Manning y H. Schütze, *Foundations of Statistical Natural Language Processing*, MIT Press: ISBN 0-262-13360-1, 1999.
- [12] H. Hewamalage, C. Bergmeir y K. Bandara, "Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions," en *International Journal of Forecasting*, vol. 37, pp. 388–427, 2020.
- [13] O. Calin, *Deep Learning Architectures*, Cham, Switzerland: Springer Nature. ISBN 978-3-030-36720-6, p. 555.
- [14] K. Greff, R. K. Srivastava, J. Koutník, *et al.*, "LSTM: A Search Space Odyssey," en *IEEE Transactions on Neural Networks and Learning Systems*, vol.28, no.10, pp. 2222–2232, 2015.

- [15] S. Hochreiter y J. Schmidhuber “Long short-term memory,” en *Neural Computation*, vol. 9, no.8, pp. 1735–1780, 1997.
- [16] Vaswani, *et al.*, “Attention is all you need,” en *NIPS’17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000–6010, 2017.
- [17] I. Solaiman, M. Brundage, J. Clark, *et al.*, *Release Strategies and the Social Impacts of Language Models*, CoRR: 1908-09203, 2019.
- [18] J. F. Kolen, *A Field Guide to Dynamical Recurrent Networks*, 1st Edition, pp. 14-24, 2001.
- [19] The Wikipedia corpus [En línea]. Disponible en: <https://www.english-corpora.org/wiki/> [Consultado en junio 02, 2021].
- [20] Commoncrawl.org, *Rastreo común*, [En línea]. Disponible en: <https://commoncrawl.org/> [Consultado en junio 02, 2021].
- [21] Radford, *et al.*, *Language Models are Unsupervised Multitask Learners*, OpenAI, 2019.

Cómo se cita:

D. E. Maqueda, “De redes neuronales recurrentes a modelos de lenguaje: la evolución del PLN en la generación de textos,” *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 4, octubre, 2021. [En línea]. Disponible en: <http://ties.unam.mx/> [Consultado en mes día, año].

Fecha de recepción: julio 03, 2021

Fecha de publicación: octubre de 2021

CLASIFICACIÓN DE POBLACIONES NATIVAS DE FRIJOL UTILIZANDO VISIÓN ARTIFICIAL

José Luis Morales Reyes, Héctor Gabriel Acosta Mesa, Elia Nora Aquino Bolaños,
Socorro Herrera Meza, Nicandro Cruz Ramírez y José Luis Chávez Servia
<http://www.ties.unam.mx/>

Fecha de recepción: julio 29, 2021 • Fecha de publicación: octubre de 2021

Octubre 2021 | número de revista 4 • ISSN en trámite



CLASIFICACIÓN DE POBLACIONES NATIVAS DE FRIJOL UTILIZANDO VISIÓN ARTIFICIAL

Resumen

El frijol es una de las fuentes más importantes de alimentación en la población mexicana. Sus propiedades colorimétricas están directamente relacionadas con sus propiedades nutrimentales. Es por eso que resulta importante contar con métodos que permitan analizar su coloración para poder comparar las características entre poblaciones. Dos de las técnicas más utilizadas para caracterizar el color en las poblaciones de frijol, han sido la espectrofotometría y el análisis de color en imágenes digitales. En el presente trabajo se propone utilizar promedios de color en los espacios de color RGB, HSI y CIE L*a*b* para representar las propiedades colorimétricas de las poblaciones. La experimentación se llevó a cabo estudiando 54 poblaciones de frijol provenientes de distintas regiones del estado de Oaxaca, México, utilizando una clase primaria (identificador de cada población) y una súper clase (formada a partir de una paleta de color). La precisión de clasificación alcanzada, utilizando el algoritmo k-NN en la clase primaria, con las mediciones promedio del espectrofotómetro, fue de 68.24%; con RGB, 46.94%; con HSI, 50%; con CIE L*a*b*, 51.48%, y conjuntando los 9 canales de los tres espacios, fue de 54.91%. Para el caso de la súper clase, los resultados obtenidos fueron de 82.87% con las mediciones promedio del espectrofotómetro, 76.2%; con RGB, 74.91%; con HSI, 77.13%; con CIE L*a*b*, y 78.8%, con los 3 espacios de color. Los resultados sugieren que aun cuando los promedios aproximan las distribuciones reales, es necesaria una representación que capture de manera más robusta las distribuciones de color asociadas a cada población.

Palabras clave:

Poblaciones Criollas de Frijol, Visión por Computadora, Aprendizaje Automático, Clasificación, Color.

CLASSIFICATION OF BEAN LANDRACES USING COMPUTER VISIÓN

Abstract

Beans are one of the most important sources of food in the Mexican population. Its colorimetric properties are directly related to its nutritional properties. That is why it is important to have methods that allow analyzing their coloration to be able to compare the characteristics between landraces. Two of the most widely used techniques to characterize color in bean landraces have been spectrophotometry and color analysis in digital images. In the present work, it is proposed to use color averages in the RGB, HSI and CIE L * a * b * color spaces to represent the colorimetric properties of the landraces. The experimentation was carried out studying 54 bean landraces from different regions of the state of Oaxaca, Mexico, using a primary class (identifier of each landrace) and a super_class (formed from a color palette). The classification precision achieved, using the k-NN algorithm in the primary class, with the average measurements of the spectrophotometer, was 68.24%; with RGB, 46.94%; with HSI, 50%; with CIE L * a * b *, 51.48%, and combining the 9 channels of the three spaces, it was 54.91%. For the case of the super_class, the results obtained were 82.87% with the average measurements of the spectrophotometer, 76.2%; with RGB, 74.91%; with HSI, 77.13%; with CIE L * a * b *, and 78.8%, with the 3 color spaces. The results suggest that even when the averages approximate the real distributions, a representation that more robustly captures the color distributions associated with each population is necessary.

Keywords:

Bean landraces, computer vision, machine learning, classification, color.

CLASIFICACIÓN DE POBLACIONES NATIVAS DE FRIJOL UTILIZANDO VISIÓN ARTIFICIAL

Introducción

El frijol y el maíz son la base de la alimentación mexicana. Ambos representan una rica fuente de proteínas, carbohidratos, vitaminas y fibra. [1] Las variedades de frijol son cultivadas a gran escala para su comercialización bajo preferencias de consumo, tipo de semilla y color. [2] Comúnmente, cada variedad está conformada por semillas con características similares, definidas por su forma y color homogéneo. Este es el criterio que permite identificar las distintas variedades comerciales. [2-4] A diferencia de una variedad comercial de frijol, una población nativa es un tipo de variedad local criolla, que ha sido preservada a través de generaciones para el consumo particular del agricultor. Estas poblaciones son el resultado de las prácticas que los agricultores realizan, esto en la selección y la mezcla de diferentes granos que se adaptan a condiciones climáticas y de suelo diversas, con lo que se ve maximizada su probabilidad de subsistencia. [1, 5-7] Las poblaciones nativas de frijol pueden estar conformadas por semillas de color homogéneo, semillas de color variegado y una mezcla de semillas de distintos patrones colorimétricos.

Las propiedades colorimétricas de una población de frijol, están directamente relacionadas con su composición química y, por ende, con sus propiedades nutrimentales. Por ello resulta importante contar con métodos, que permitan analizar su color para el estudio de las re-

laciones que existen entre las propiedades colorimétricas y las químicas, además de comparar las similitudes y las diferencias entre las poblaciones.

Una de las técnicas más utilizadas para caracterizar el color en las poblaciones de frijol, es la espectrofotometría. [8, 9] Esto se lleva a cabo con un equipo de análisis de color de alta precisión, llamado espectrofotómetro. [10, 11] Éste cuenta con una abertura de 8 mm, donde es colocado el objeto a medir. Dado que las mediciones con un espectrofotómetro son puntuales, medir la totalidad de color de un grano no es práctico, por lo que varias mediciones de color son realizadas sobre la semilla. [8, 12] La dificultad para representar el color de un grano aumenta cuando está conformado por diferentes tonalidades (variegados). Una alternativa de caracterización del color es la que se hace mediante un sistema de visión artificial, a través del cual es posible capturar el color de una imagen adquirida por medio de una cámara digital. [13]

A continuación se reportan los trabajos que hacen uso de las propiedades colorimétricas de los granos. En *Digital camera images processing of hard-to-cook beans*, [14] se representa el color de los granos mediante 4 valores obtenidos de los histogramas: el valor del nivel mínimo y máximo del rango dinámico, el valor del nivel con mayor frecuencia y el valor de la frecuencia. En el estudio *Identification of bean varieties according to color features using artificial neural network*, [15] las caracterís-

tics colorimétricas son representadas por el promedio del color principal y el promedio de las manchas de la semilla, en caso de contenerlas.

En el trabajo *Beans quality inspection using correlation-based granulometry*, [16] se considera como característica el color dominante que presenta una semilla, el cual comprende el 70% de los píxeles. Algunos trabajos no se limitan a trabajar únicamente con color. Se emplean descriptores de textura y forma, considerados para la clasificación; [17, 18] sin embargo, debido a que nuestro interés es estudiar la relación entre las propiedades colorimétricas y las nutrimentales, nuestro análisis se focaliza solamente en el color.

En la literatura existen diversos trabajos de clasificación de semillas. En muchos de ellos se reporta la clasificación individual de éstas, mientras que en el presente trabajo, la población de frijol se analiza como un conjunto de semillas que conforma la población. Se propone la utilización de un sistema de visión artificial, cuya viabilidad se evalúa mediante el cálculo del índice de precisión al clasificar las poblaciones. Como parte de la validación del método, se comparan los resultados de la precisión de clasificación de nuestra propuesta, contra las obtenidas mediante mediciones puntuales aleatorias, adquiridas con el espectrofotómetro.

El artículo está dividido en cuatro secciones: en la primera se exponen los materiales y los métodos utilizados en el proyecto. La segunda muestra los resultados obtenidos con nuestra propuesta de representación. En la tercera sección se discuten los resultados, para, finalmente, en la cuarta dar las conclusiones y exponer el trabajo futuro.

Materiales y métodos

Las poblaciones nativas de frijol, utilizadas en el presente proyecto, proceden de distintas regiones del estado de Oaxaca. Todas ellas pertenecen a variedades locales criollas, donde el cultivo es la fuente principal de alimento. En este trabajo se consideran 54 poblaciones de frijol. Cada una es representada con una muestra de 20 granos. En la figura 1 se muestran algunos ejemplos de poblaciones incluidas en este trabajo.

Entorno de iluminación

El prototipo utilizado en este trabajo, tiene la finalidad de ofrecer un entorno de iluminación controlada y de intensidad uniforme. Fue fabricado a base de

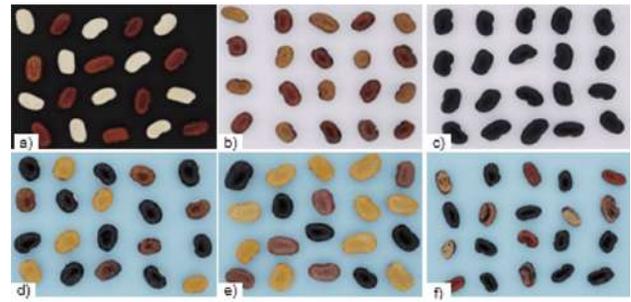


Figura 1. Ejemplos de poblaciones de frijol. De izquierda a derecha: a) Mezcla negro y rojo, b) Mezcla café y rojo, c) Negro, d) Mezcla amarillo, amarillo con café y negro, e) Mezcla amarillo, negro y rosa, f) Mezcla amarillo con negro, café, negro, rojo y rosa con negro

triplay, con medidas de 30x30x45 cm de largo, ancho y alto, respectivamente. La conexión eléctrica utilizada en el interior, fue en paralelo. Se utilizaron cuatro fuentes luminosas de 40 watts de luz cálida, que fueron colocadas en cada una de las esquinas, para proporcionar intensidad y dispersión de luz uniforme (figura 2a).

Caja de difuminación

En este trabajo se empleó una caja de difuminación como complemento del entorno de iluminación, con la finalidad de reducir reflejos especulares y evitar sombras entre granos. La caja de luz es un cubo con dimensiones 19.5x19.5x19.5 cm de largo, ancho y alto, respectivamente. Fue forrado con tela translúcida en cuatro laterales, dejando libre la parte superior para la visibilidad y la adquisición de imagen. La figura 2b muestra la caja de luz empleada en este trabajo.

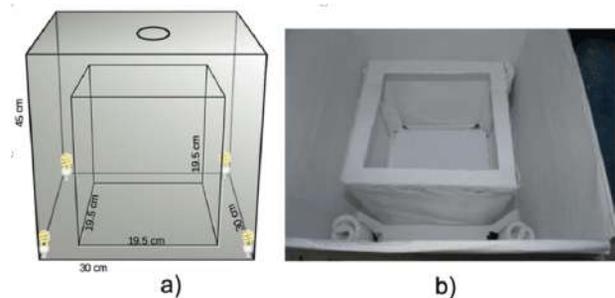


Figura 2. a) Entorno de iluminación, b) Caja de difuminación en el interior del prototipo de iluminación.

Utilizando este dispositivo, la difuminación de luz prácticamente eliminó los reflejos y las sombras. La intensidad luminosa en el interior del prototipo de iluminación fue de 3600 Lx. Fue medida con un fotómetro sekonic multimeter l-408.

Cámara fotográfica digital

En este trabajo se utilizó una cámara fotográfica digital SONY, modelo ILCE 3500 de 20 MP, con configuración de disparo mostrada en la tabla 1. La configuración sirvió para estandarizar la adquisición de imágenes. Adicionalmente, la corrección automática de brillo y contraste fue deshabilitada. El balance de blancos se indicó incandescente, como especificación al tipo de fuente luminosa. Se utilizó un zoom óptico de 50 mm para la adquisición de imagen a una distancia de 45 cm del objetivo.

El proceso de adquisición de las imágenes consistió en colocar dentro de la caja de difuminación una muestra de 20 granos, separados entre sí, para evitar oclusiones y/o sombras entre granos. El fondo de color negro se utilizó para contrastar las poblaciones con granos de color blanco. El fondo blanco se usó para contrastar poblaciones de color negro y rojo, así como combinar ambas. Por último, el fondo de color azul se utilizó para contrastar las poblaciones con granos de distintos colores.

| Opción de configuración | | Opción de configuración | |
|-------------------------|---------------|-------------------------|----------|
| Modo | Manual | Exposición | 1/50 |
| Apertura | 5.6 | Zoom óptico | 50 mm |
| Enfoque | Manual | ISO | 100 |
| Balance blanco | Incandescente | Área de enfoque | Múltiple |

Tabla 1. Configuración de disparo de la cámara fotográfica

Segmentación de imágenes

Como primer paso, fue necesario identificar en cada imagen la superficie correspondiente a las semillas. Para esto se utilizó el algoritmo de segmentación por crecimiento de regiones. Este algoritmo busca agrupar píxeles, con base en un criterio de similitud predefinido. [19, 20]

Para el presente trabajo, el criterio de similitud fue definido por el atributo de color. Los tres colores utilizados como fondos de contraste, se usaron como atributo predefinido para la segmentación. Como métrica de similitud entre píxeles, se utilizó la distancia euclidiana ponderada en el espacio de color CIE L*a*b*, por ser el que mejores resultados arrojó.

$$D_{lab} = \sqrt{(w_1 * l)^2 + (w_2 * a)^2 + (w_3 * b)^2} \quad (1)$$

Para la obtención del valor de los pesos w_i de cada componente de color, se aplicó la fórmula:

$$w_i = 1/\sigma_i \quad (2)$$

Donde:

w_i : es el resultado de la normalización de cada componente de color

σ_i : es la desviación estándar del componente de color

Caracterización del color

Una imagen digital puede ser representada por una función de dos dimensiones, cuyas coordenadas y valores de amplitud son enteros. El número de niveles de gris (k), comúnmente está expresado como una potencia de 2n. [19] Un espacio de color está conformado por tres canales. Por lo tanto, es posible obtener el promedio de los valores de cada canal y conformar el vector de tres valores. Los espacios de color considerados, fueron RGB (Rojo, Verde, Azul), HSI (Matiz, Saturación, Intensidad) y CIE L*a*b*, por ser los más comúnmente utilizados en trabajos relacionados. [14-18]

Paleta de colores

El objetivo principal del presente trabajo es la clasificación de las poblaciones de frijol, con respecto a su color. La clase principal está definida por la etiqueta que identifica a cada población (clase primaria); sin embargo, con la finalidad de estudiar cómo se agrupan estas poblaciones, de acuerdo a sus similitudes colorimétricas, fue creada una paleta de colores (súper clase), que agrupa las semillas de acuerdo a la coloración del grano. Esta paleta de colores contiene doce categorías (figura 3). La relación entre las clases primarias y la súper clase, se detalla en la tabla 2.

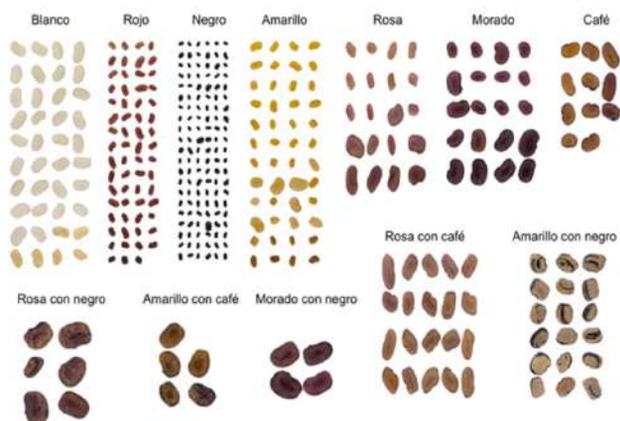


Figura 3. Paleta de color creada con los granos de las poblaciones de frijol. Fue utilizada como referencia para generar una súper clase por población de frijol.

Aprendizaje automático

Aun cuando se experimentó con diferentes algoritmos de clasificación, como máquinas de soporte vectorial y árboles de decisión, la clasificación de las poblaciones de frijol se realizó con el algoritmo de aprendizaje supervisado k vecinos más cercanos o k -NN (k -Nearest Neighbor), por haber obtenido mejores resultados. Para la clasificación, se utilizaron diferentes métricas de similitud, encontrándose los mejores resultados con la métrica Cityblock. Se utilizó la distancia cuadrada inversa, para ponderar el peso de cada uno de los k vecinos, de acuerdo a su proximidad a la instancia a clasificar. [21]

Los resultados de clasificación fueron medidos en términos de precisión, que se calculó a partir del promedio de las instancias correctamente clasificadas. Como método de evaluación, se utilizó la validación cruzada, particularmente la técnica de holdout, considerándose 50% de los datos para entrenamiento y el restante 50%, para validación. Se ejecutaron 20 corridas. Los algoritmos de procesamiento digital de imagen y aprendizaje automático, se implementaron en Matlab 2019b, y fueron ejecutados en una estación de trabajo Dell Precision T3610.

Metodología

La metodología propuesta para el análisis de clasificación, se muestra en la figura 4. En primer término, se utilizan las imágenes de cada población para realizar la separación

| No. | Clase primaria | Súper Clase | No. | Clase primaria | Súper Clase |
|-----|----------------|---|-----|----------------|-------------------------------|
| 1 | Pob-02 | Blanco | 28 | Pob-43 | Mezcla amarillo y rojo |
| 2 | Pob-03 | Mezcla negro y rojo | 29 | Pob-44 | Negro |
| 3 | Pob-04 | Negro | 30 | Pob-46 | Negro |
| 4 | Pob-05 | Blanco | 31 | Pob-48 | Negro |
| 5 | Pob-06 | Negro | 32 | Pob-49 | Negro |
| 6 | Pob-08 | Negro | 33 | Pob-50 | Blanco |
| 7 | Pob-09 | Rojo | 34 | Pob-51 | Mezcla amarillo, negro y rojo |
| 8 | Pob-11 | Amarillo | 35 | Pob-52 | Negro |
| 9 | Pob-13 | Negro | 36 | Pob-54 | Mezcla negro, rojo y rosa |
| 10 | Pob-17 | Mezcla blanco y rojo | 37 | Pob-55 | Negro |
| 11 | Pob-18 | Negro | 38 | Pob-56 | Mezcla blanco y negro |
| 12 | Pob-19 | Rojo | 39 | Pob-57 | Amarillo |
| 13 | Pob-20 | Mezcla amarillo, morado, morado con negro, negro y rosa | 40 | Pob-58 | Rojo |
| 14 | Pob-22 | Mezcla morado, morado con negro, rosa y rosa con negro | 41 | Pob-59 | Amarillo |
| 15 | Pob-23 | Mezcla amarillo con negro, café, negro, rojo y rosa con negro | 42 | Pob-61 | Negro |
| 16 | Pob-26 | Mezcla amarillo, negro y rosa | 43 | Pob-62 | Rojo |
| 17 | Pob-28 | Mezcla negro y rojo | 44 | Pob-63 | Rojo |
| 18 | Pob-29 | Negro | 45 | Pob-64 | Rojo |
| 19 | Pob-30 | Negro | 46 | Pob-65 | Amarillo |
| 20 | Pob-31 | Mezcla morado y negro | 47 | Pob-66 | Mezcla negro y rojo |
| 21 | Pob-32 | Mezcla amarillo, amarillo con café y negro | 48 | Pob-67 | Blanco |
| 22 | Pob-33 | Mezcla amarillo, negro y rosa | 49 | Pob-68 | Amarillo con negro |
| 23 | Pob-34 | Mezcla amarillo y amarillo con negro | 50 | Pob-69 | Mezcla morado y negro |
| 24 | Pob-35 | Mezcla morado y negro | 51 | Pob-71 | Rosa con café |
| 25 | Pob-39 | Mezcla café y rojo | 52 | Pob-72 | Mezcla amarillo, negro y rojo |
| 26 | Pob-40 | Rojo | 53 | Pob-74 | Mezcla blanco, negro y rojo |
| 27 | Pob-41 | Negro | 54 | Pob-76 | Rojo |

Tabla 2. Lista de etiquetas de clase, utilizadas para la clasificación de poblaciones nativas de frijol. La clase fue utilizada para clasificar poblaciones por variedad local y la súper clase fue utilizada para clasificar por coloración las semillas

aleatoria de las semillas, con la finalidad de formar conjuntos de entrenamiento y prueba de cada población de frijol. El siguiente paso consistió en la conversión del espacio de color RGB a HSI y CIE L*a*b*, para posteriormente extraer las características colorimétricas de las semillas representadas como promedios. Estos promedios fueron calculados por cada canal, de cada uno de los 3 espacios de color, obteniéndose un vector de 3 características para cada uno de ellos. Finalmente, se aplicó el algoritmo de aprendizaje supervisado k -NN. Los resultados de las precisiones reportadas, se calcularon como los promedios de realizar 20 corridas de este proceso.

Resultados

Con la finalidad de comparar los resultados de clasificación obtenidos con nuestra propuesta, contra los obtenidos por un método tradicional, como la adquisición colorimétrica puntual, adquirida con un espectrofotómetro, se procedió a realizar la medición de color con un espectrofotómetro en cada una de las poblaciones utilizadas en este trabajo. [22] Por otra parte, fue calculado el vector de promedios de cada canal de los espacios de color RGB; HSI y CIE L*a*b*, obtenidos de las regiones de interés. Cabe aclarar que aun cuando se realizaron pruebas con distintos valores de k en el algoritmo k -vecinos cercanos, por cuestiones de espacio solo se presentan los casos donde se obtuvieron los mejores resultados. En la tabla 3 se muestran los resultados de precisión de clasificación, para la clase primaria (población).

En la tabla 4 se muestran los resultados en precisión de clasificación para la súper clase (paleta de color).

En la figura 5 se muestran las matrices de confusión, pertenecientes a los resultados obtenidos con el vector compuesto por los 9 valores, correspondientes a los 3 espacios de color, el mismo con que se obtuvo el mejor resultado en el análisis de imagen.

| Método | Media y desviación estándar |
|------------------------------|-----------------------------|
| Espectrofotómetro (k=1) | 68.24±4.59 |
| HSI (k=1) | 50.00±5.70 |
| CIE L*a*b* (k=1) | 51.48±6.43 |
| RGB (k=1) | 46.94±6.78 |
| HSI + CIE L*a*b* + RGB (k=1) | 54.91±6.28 |

Tabla 3. Lista de etiquetas de clase utilizadas para la clasificación de poblaciones nativas de frijol. La clase fue utilizada para clasificar poblaciones por variedad local y la súper clase fue utilizada para clasificar por coloración de las semillas.

| Método | Media y desviación estándar |
|------------------------------|-----------------------------|
| Espectrofotómetro (k=5) | 82.87±3.98 |
| HSI (k=3) | 74.91±3.03 |
| CIE L*a*b* (k=1) | 77.13±4.78 |
| RGB (k=1) | 76.20±5.42 |
| HSI + CIE L*a*b* + RGB (k=3) | 78.80±4.52 |

Tabla 4. Resultados en precisión de clasificación. Valores reportados como media ± desviación estándar de 20 corridas. Se especifica el valor k , con el que se obtuvieron las precisiones más altas.

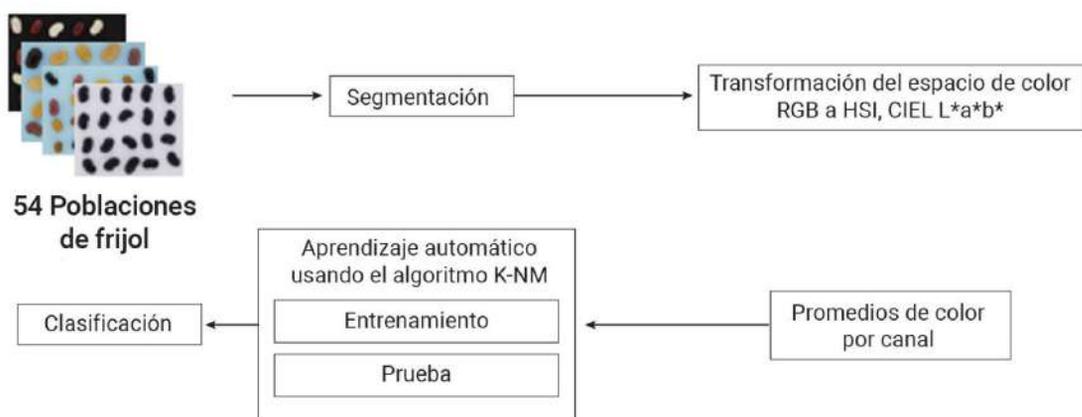
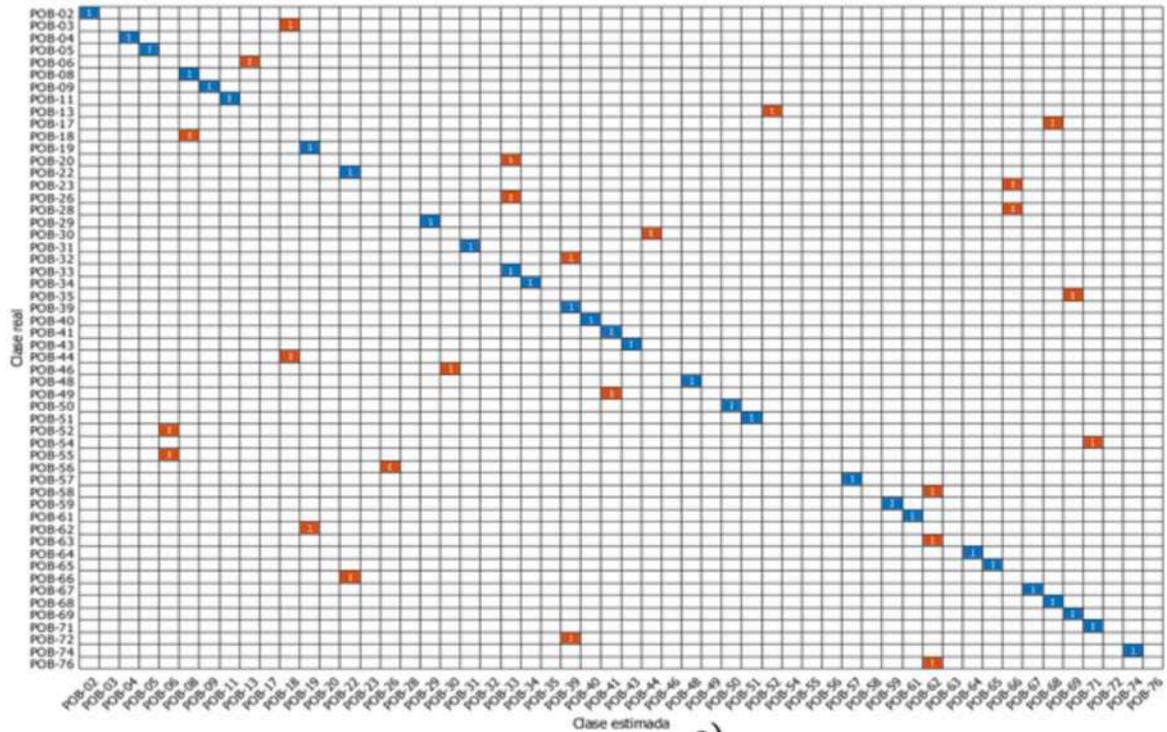
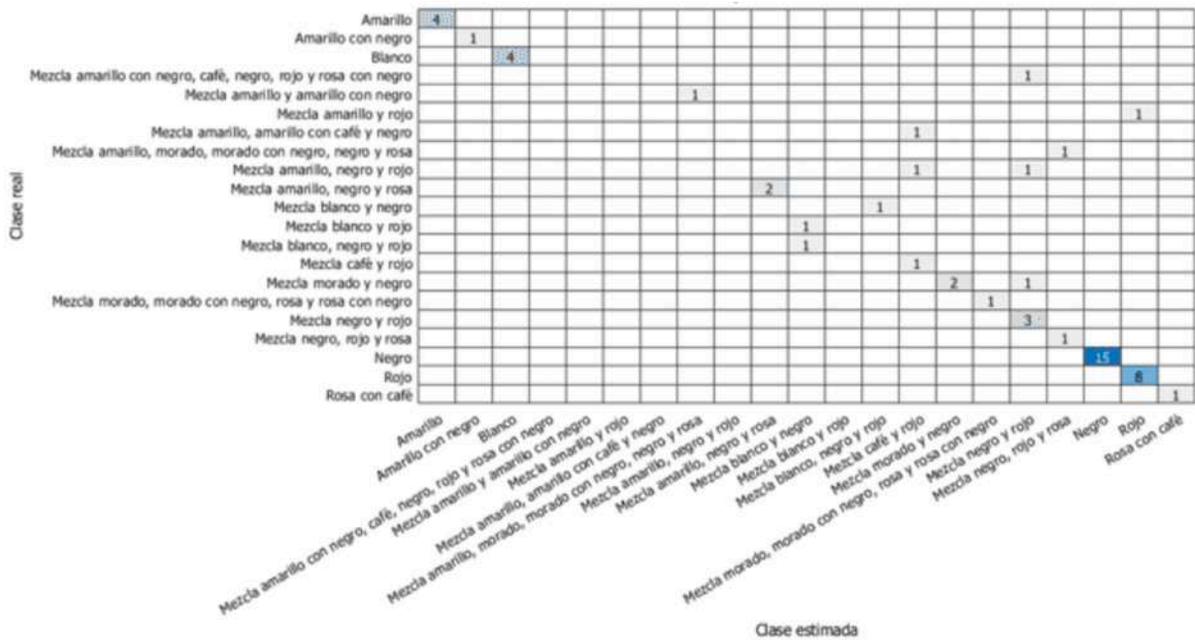


Figura 4. Procedimiento para la generación de resultados de clasificación de poblaciones nativas de frijol



a)



b)

Figura 5. Matrices de confusión. a) Clase. Las letras entre paréntesis hacen referencia al color predominante en la población: (A) Amarillo, (B) Blanco, (M) Mezcla de granos de diferente color, (N) Negro, (R) Rojo
b) Súper clase

Discusión

La precisión de clasificación alcanzada con las mediciones promedio del espectrofotómetro, HSI, CIE $L^*a^*b^*$, RGB y conjunto de HSI+CIELab+RGB, fueron 68.24 %, 50.0% , 51.48%, 46.94% y 54.91%, respectivamente. Lo que se muestra es que aun cuando el método propuesto alcanza un nivel de precisión cercano al del espectrofotómetro, no lo logra superar. Esto sugiere que los nueve promedios no son suficientes para capturar las diferencias entre las poblaciones. Un comportamiento similar se observó en los resultados de clasificación obtenidos, usando la etiqueta súper clase. Las matrices de confusión muestran que los errores de clasificación para el caso Clase, se deben a que existen diferentes poblaciones con colores similares, cuyos valores no pueden ser discriminados a partir de los promedios. Tal es el caso de la POB-58 roja, que fue clasificada como POB-62 roja (ver figura 5a). De manera similar, para el caso de las poblaciones heterogéneas o mezcla de granos de varios colores, presentan una problemática similar al clasificar erróneamente poblaciones que comparten algunos de los colores (ver figura 5b).

Conclusiones

Las condiciones de iluminación controladas y la configuración de la cámara, fueron determinantes para estandarizar la adquisición fotográfica, permitiendo la reproducibilidad del método. Trabajos relacionados utilizan promedios de color de granos. Además, muchos de ellos se enfocan en la clasificación individual de granos y no en su población en conjunto. En este trabajo se mostró que el uso de los promedios de color fue suficiente para alcanzar niveles de precisión, similares a los obtenidos con un espectrofotómetro. Debido a que las poblaciones de frijol son variedades que pueden estar conformadas por semillas similares o por una combinación de semillas de distintos colores, como trabajo futuro, se propone utilizar una forma de representación de color que considere el uso de la tonalidad total de una semilla, por ejemplo, histogramas de cromaticidad que capturen toda la distribución de probabilidad de los colores y no solo sus medias. Esto podría mejorar considerablemente los niveles de precisión de nuestro método.

Reconocimientos

El primer autor de este artículo agradece al Consejo Nacional de Ciencia y Tecnología (CONACyT), la beca (712056) concedida para la realización de estudios doctorales.

BIBLIOGRAFÍA

- [1] S. H. Guzman-Maldonado and O. Paredes-López, "Pérdidas irreversibles de un rico legado de la naturaleza para la humanidad. El caso del frijol común en México," *Ciencia*, vol. 49, no. 4, pp. 5-13, 1998.
- [2] P. Graham and P. Ranalli, "Common bean (*Phaseolus vulgaris* L.)," *Field Crops Research*, vol. 53, no. 1-3, pp. 131-146, 1997.
- [3] D. L. Luthria and M. A. Pastor-Corrales, "Phenolic acids content of fifteen dry edible bean (*Phaseolus vulgaris* L.) varieties," *Journal of Food Composition and Analysis*, vol. 19, no. 2, pp. 205-211, 2006/03/01/ 2006, doi: <https://doi.org/10.1016/j.jfca.2005.09.003>.
- [4] C. W. a. H. Beninger, L. George, "Antioxidant Activity of Extracts, Condensed Tannin Fractions, and Pure Flavonoids from *Phaseolus vulgaris* L. Seed Coat Color Genotypes," *Journal of Agricultural and Food Chemistry*, vol. 51, pp. 7879--7883, 2003/12/01 2003, doi: 10.1021/jf0304324.
- [5] J. L. Chávez-Servia, *et al.*, "Diversity of common bean (*Phaseolus vulgaris* L.) landraces and the nutritional value of their grains," in *Grain Legumes: InTech*, 2016.
- [6] Y. D. García-Díaz, E. N. Aquino-Bolaños, J. L. Chávez-Servia, *et al.*, "Bioactive compounds and antioxidant activity in the common bean are influenced by cropping season and genotype," *Chilean journal of agricultural research*, vol. 78, no. 2, pp. 255-265, 2018.
- [7] A. Capistrán-Carabarin, E. N. Aquino-Bolaños, Y. D. García-Díaz, *et al.*, "Complementarity in Phenolic Compounds and the Antioxidant Activities of *Phaseolus coccineus* L. and *P. vulgaris* L. Landraces," *Foods*, vol. 8, no. 8, p. 295, 2019.
- [8] E. N. Aquino-Bolaños, Y. D. García-Díaz, J. L. Chavez-Servia, *et al.*, "Anthocyanins, polyphenols, flavonoids and antioxidant activity in common bean (*Phaseolus vulgaris* L.) landraces," *Emirates Journal of Food and Agriculture*, pp. 581-588, 2016.
- [9] C. Montalvo-Paquini, R. Avila-Sosa, A. López-Malo, *et al.*, "Preparation and Characterization of Proteinaceous Films from Seven Mexican Common Beans (*Phaseolus vulgaris* L.)," *Journal of Food Quality*, vol. 2018, p. 9782591, 2018/01/03 2018, doi: 10.1155/2018/9782591.
- [10] A. H. Beckett and J. B. Stenlake, "Practical Pharmaceutical Chemistry: Part II Fourth Edition," *Bloomsbury Academic*, Practical Pharmaceutical Chemistry, pp. 255-272, 1988. [Online]. Available: <https://books.google.com.mx/books?id=Up3L2dA17k8C>. [Accessed february 14, 2019]

- [11] R. D. Simoni, R. L. Hill, M. Vaughan, *et al.*, "A classic instrument: The Beckman DU spectrophotometer and its inventor, Arnold O. Beckman," *Journal of Biological Chemistry*, vol. 278, pp. 79–81, 2003, doi: doi:10.1016/s0021-9258(20)75750-9.
- [12] L. G. Espinosa-Alonso, A. Lygin, J. M. Widholm, *et al.*, "Polyphenols in wild and weedy Mexican common beans (*Phaseolus vulgaris* L.)," *Journal of agricultural and food chemistry*, vol. 54, no. 12, pp. 4436-4444, 2006, doi: 10.1021/jf060185e.
- [13] J. C. del Valle, A. Gallardo-López, M. L. Buide, *et al.*, "Digital photography provides a fast, reliable, and noninvasive method to estimate anthocyanin pigment concentration in reproductive and vegetative plant tissues," vol. 8, no. 6, pp. 3064-3076, 2018, doi: doi:10.1002/ece3.3804.
- [14] B. Laurent, B. Ousman, T. Dzudie, *et al.*, "Digital camera images processing of hard-to-cook beans," *Journal of Engineering and Technology Research*, vol. 2, no. 9, pp. 177-188, 2010.
- [15] A. Nasirahmadi and N. Behroozi-Khazaei, "Identification of bean varieties according to color features using artificial neural network," *Spanish Journal of Agricultural Research*, vol. 11, no. 3, pp. 670-677, 2013.
- [16] S. A. De Araújo, J. H. Pessota, and H. Y. Kim, "Beans quality inspection using correlation-based granulometry," *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 84-94, 2015.
- [17] M. L. Bianco, O. Grillo, R. Cremonini, *et al.*, "Characterisation of Italian bean landraces ('*Phaseolus vulgaris*' L.) using seed image analysis and texture descriptors," *Australian Journal of Crop Science*, vol. 9, no. 11, p. 1022, 2015.
- [18] G. Venora, O. Grillo, C. Ravalli, *et al.*, "Identification of Italian landraces of bean (*Phaseolus vulgaris* L.) using an image analysis system," *Scientia Horticulturae*, vol. 121, no. 4, pp. 410-418, 2009.
- [19] R. C. Gonzalez and R. E. Woods, "Digital image processing," 2002.
- [20] T. Jun, "A color image segmentation algorithm based on region growing," in *2010 2nd International Conference on Computer Engineering and Technology*, 16-18 April 2010 2010, vol. 6, pp. V6-634-V6-637, doi: 10.1109/ICCET.2010.5486012.
- [21] G. Batista and D. F. Silva, "How k-nearest neighbor parameters affect its performance," in *Argentine symposium on artificial intelligence*, 2009: sn, pp. 1-12.
- [22] R. G. McGuire, "Reporting of objective color measurements," *HortScience*, vol. 27, no. 12, pp. 1254-1255, 1992.

Cómo se cita:

J. L. Morales, H. G. Acosta, E. N. Aquino, *et al.*, "Clasificación de poblaciones nativas de frijol utilizando visión artificial," *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 4, octubre, 2021. [En línea]. Disponible en: <http://ties.unam.mx/> [Consultado en mes día, año].

Fecha de recepción: julio 29, 2021

Fecha de publicación: octubre de 2021

LAS ANOMALÍAS: ¿QUÉ SON?, ¿DÓNDE SURGEN?, ¿CÓMO DETECTARLAS?

Nidiyare Hevia Montiel , Sergio Mota y Antonio Neme
<http://www.ties.unam.mx/>

Fecha de recepción: agosto 2, 2021 • Fecha de publicación: octubre de 2021

Octubre 2021 | número de revista 4 • ISSN en trámite



LAS ANOMALÍAS: ¿QUÉ SON?, ¿DÓNDE SURGEN?, ¿CÓMO DETECTARLAS?

Resumen

Una anomalía es una instancia que no se asemeja a la mayoría de las observaciones. Establecer los criterios de comparación, que nos permitan identificar a una instancia como posible anomalía es una tarea abierta en Inteligencia Artificial (IA). La baja frecuencia de estas dificulta tener datos para extraer atributos, que nos proporcionen una idea de qué hace diferente a una anomalía con respecto a las observaciones usuales o habituales. La idea general de los algoritmos de detección de anomalías pasa por escudriñar las observaciones usuales o habituales, para extraer algún criterio o métrica que sea compartida por ellas, y que posea la propiedad de no estar presente en observaciones anómalas. La práctica tradicional para detectarlas es asociarlas a ruido o error en las observaciones y descartarlas. Una perspectiva moderna dice que: una anomalía o discrepancia es un indicio, posiblemente temprano, de algún cambio importante en el objeto de estudio. En este trabajo, presentaremos definiciones operativas, relataremos en qué contextos surgen, y haremos un recorrido sobre algunos algoritmos para su detección.

Palabras clave:

Aprendizaje no supervisado, detección de anomalías, densidad, distancias.

ANOMALIES: WHAT ARE THEY? WHERE DO THEY ARISE? HOW TO DETECT THEM?

Abstract

An anomaly is an instance that does not resemble most observations. Establishing the comparison criteria that allow us to identify an instance as a possible anomaly is an open task in Artificial Intelligence. Anomalies are generally much less frequent than common or usual observations. The low frequency of occurrence of the candidate anomalies makes it difficult to have data sets from which it can extract attributes that give us an idea about what makes an anomaly different from the common or usual observations. The core idea of anomaly detection algorithms is to scrutinize the common or usual observations and extract a criterion or metric that is common among those observations, at the time that such property is absent in the anomalous observations. Traditional practice with respect to anomalies is to associate them with observations noise or errors, and then discard them. A modern perspective suggests us that an anomaly is an indication, possibly an early one, of some major change in the studied phenomena. In this paper, we will present operational definitions of anomalies, we will relate in which contexts they arise, and we will take a tour of some algorithms for their detection.

Keywords:

Anomaly detection, Density, Distances, Unsupervised learning.

LAS ANOMALÍAS: ¿QUÉ SON?, ¿DÓNDE SURGEN?, ¿CÓMO DETECTARLAS?

Introducción

Imagine el siguiente escenario. Usted, como todos los días en los últimos años, sale de su casa más o menos a la misma hora, y mientras camina a la estación del metro o a la parada del autobús, observa que algo parece diferente a lo que cotidianamente percibe. Le lleva un momento observar, finalmente, ubica la fuente de lo extraño: el puesto de *guajolotas*¹ se movió unos metros hacia la esquina. Ahora, imagine un segundo escenario. Sale de su casa, y en su caminata a la entrada de la estación del metro, se da cuenta que algo es diferente. Le lleva unos segundos percatarse de lo que se trata: hay un nuevo puesto de tacos en la esquina.

¿Cuál de los dos escenarios es una anomalía? ¿Qué aspectos de la escena visual compara con sus recuerdos, el primer caso es apenas una variación de lo que cotidianamente observa, en tanto que, en el segundo hay algo cualitativamente distinto? Responder a estas preguntas es parte de la problemática a la que se enfrentan los algoritmos de detección de anomalías.

Para llevar al lector a un ámbito más concreto, se le pide observar la figura 1(a), en la que se muestra una distribución de $N = 19$ puntos, donde dos constituyen anomalías o puntos atípicos. De acuerdo con criterios que

describiremos más adelante. Se invita al lector a observar dicha figura con atención antes de seguir con la lectura. ¿Puede usted identificar cuáles serían los puntos candidatos a ser considerados atípicos? ¿Qué lo llevó a hacer su elección de considerar esos puntos como anómalos?

Una de las caracterizaciones más comunes que nos permiten darnos una idea de la diversidad de los datos y, con ello, una aproximación a su nivel de anomalía es la de *distancia*. Podemos pensarla como una función que permite comparar cosas; mientras más semejantes sean dos objetos, menor es su distancia. De esta forma, entonces, podemos caracterizar la distancia entre cada punto y su vecino más cercano. La figura 1(b) muestra el diámetro que representa a cada círculo en función de la distancia con el vecino más próximo. El vector C es representado por el círculo más grande, puesto que la distancia a su punto más cercano, L , es muy alta. Los puntos R y S son representados por círculos pequeños, porque la distancia entre ellos es baja.

Para seguir en esa línea, podemos obtener la distancia al segundo punto más próximo. De esta forma, contaremos con dos caracterizaciones o atributos para cada vector: la distancia a su primer y segundo vecino.

Los 19 puntos se grafican en este nuevo espacio de atributos, como lo muestra la figura 1(c). En esta nueva caracterización, se observa que el vector C es muy diferente a los demás, pues se encuentra en una zona muy alejada del resto de los puntos, es decir, aislado. El

¹ Entiéndase por una torta de tamal. Véase: <https://es.wikipedia.org/wiki/Guajolota>

punto C sería considerado entonces, bajo estos criterios descritos, una anomalía. En la figura 1(d) se muestra a cada punto rodeado por una región circular del mismo radio. Una caracterización de esta región, o *vecindario*, pasa por contar el número de vectores o puntos dentro de ella. Por lo que, bajo este criterio observamos que 17 de 19 puntos cuentan con dos o tres vecinos, con excepción de los puntos *c* y *f* donde el primero carece de vecinos, en tanto que el segundo cuenta con cuatro de ellos.

Las anomalías son eventos que no se asemejan a los eventos usuales [1-3] y surgen en una diversidad de ámbitos. Estas se presentan dentro del área de la medicina, en donde formaciones tumorales son anomalías con respecto a tejido sano; [4] se presentan en biología molecular, cuando ciertos genes se expresan en situaciones en las que no se esperaría que lo hicieran; [5] en las artes, cambios en el estilo de escritura de un autor se identifican con anomalías causadas probablemente por algún trastorno cognitivo. [6,7]

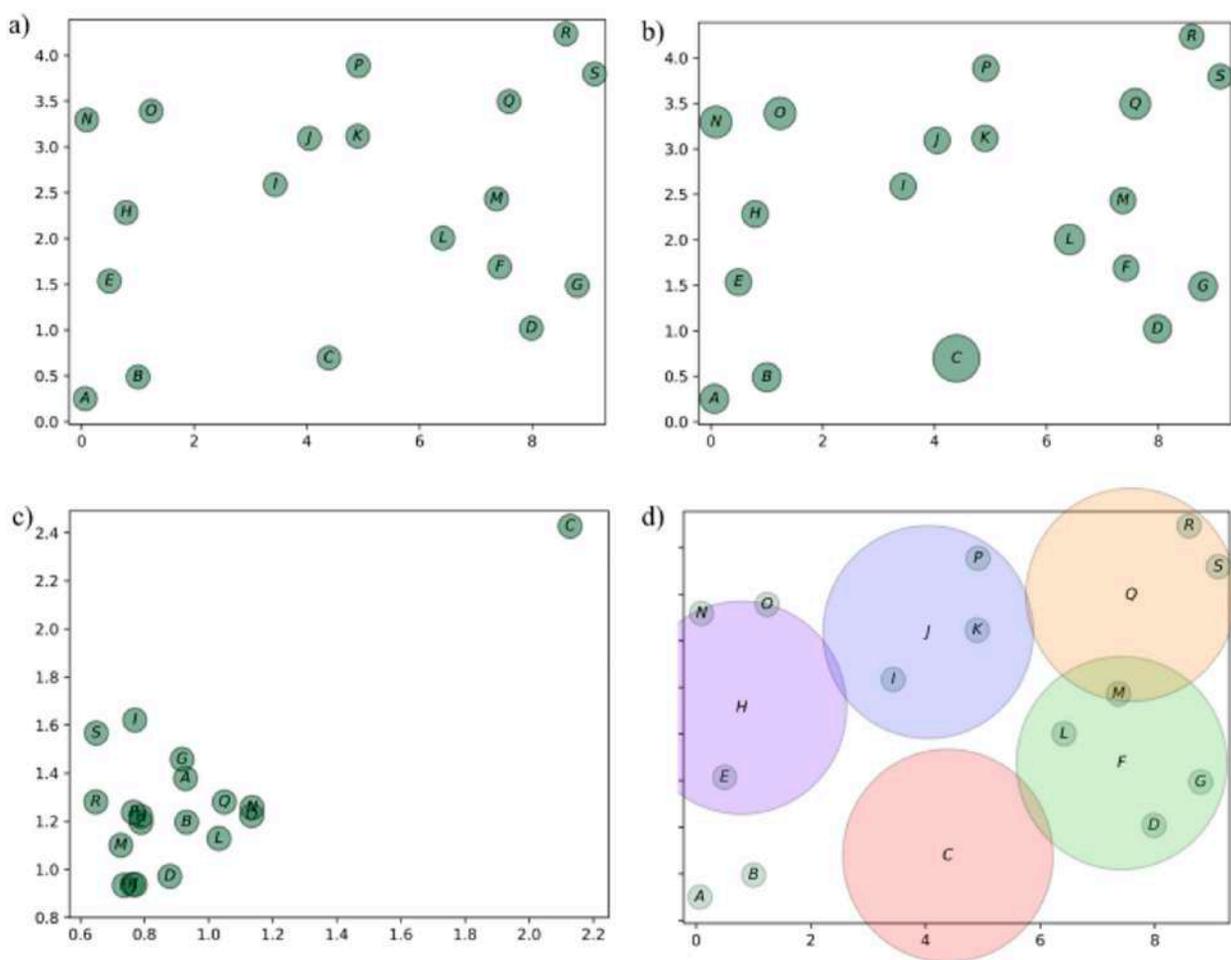


Figura 1. $N = 19$ puntos. (a): ¿Existe alguna propiedad que sea compartida por la mayoría de los puntos, y que no sea satisfecha por un número reducido de ellos? (b): El radio de cada punto es una función de la distancia al vecino más próximo. El punto C es el más aislado de entre todos los puntos, por lo que el círculo asociado a éste es el de mayor radio. (c) El eje x muestra la distancia al punto más cercano, en tanto que el eje y muestra la distancia al segundo vecino más próximo; por lo que el punto C se muestra claramente distinto a los demás. (d) Se define un vecindario o vecindad mediante un círculo para cada punto, donde aquellos que se encuentren dentro del círculo serán los vecinos. El punto J cuenta con I, K y P como vecinos. Se muestra únicamente el vecindario de los puntos C, F, H, J y O.

Se le conoce como *detección de anomalías* al proceso de identificación de observaciones posiblemente anómalas. [8,9] Los algoritmos para detectarlas caen dentro de una rama de la Inteligencia Artificial (IA) conocida como aprendizaje no supervisado. La IA es una disciplina cobijada dentro de las Ciencias de la Computación, pero es resultado de la interacción de investigadores y profesionales de muy diversas disciplinas. [10] Dentro de la IA, un campo de particular interés es el aprendizaje automático, [11] que responde a preguntas tales como: ¿Es posible que un algoritmo aprenda a hacer algo sin que explícitamente se le enseñe la manera de hacerlo? ¿Es posible encontrar un algoritmo que aprenda a distinguir la clase o categoría a la que alguna instancia pertenece?

Los éxitos de la IA en el aprendizaje supervisado han sido notables en muchos ámbitos. Este aprendizaje se da cuando cada punto lleva asociada una etiqueta o clase. Por ejemplo, a una imagen médica, como podría ser el caso de una resonancia magnética cerebral, se le puede asignar la etiqueta de *tejido sano* o *tejido con tumor*. Para este caso, un sistema de aprendizaje supervisado intentará encontrar implícitamente una función que asocie a cada píxel de la imagen, o su descripción en realidad, con la clase o etiqueta asociada. De esta forma, el sistema aprenderá a definir en una imagen lo que representa tejido sano y qué lo distingue como tejido tumoral. Este sistema, al ser interrogado sobre una imagen que nunca había analizado, es decir, que no estaba presente en el conjunto de entrenamiento, deberá de decidir cuáles píxeles pertenecen a la clase de tejido sano o de tumor.

Desarrollo

Una pregunta pertinente dentro de la IA es la siguiente: ¿Puede un sistema aprender a identificar situaciones inéditas sin que explícitamente un supervisor le diga cuando se encuentra en presencia de una? El proceso de clasificación requiere de una muestra representativa y relevante de las clases de interés. En el caso de detección de las anomalías, la clase correspondiente es extremadamente infrecuente, y por ello, poco numerosa. Por otro lado, la clase de las instancias usuales o habituales suele ser mucho más frecuente que la primera. Esto podría estudiarse bajo la perspectiva de lo que en IA se conoce como *clasificación con clases desbalanceadas*. [8]. Esto significa que hay dos grupos: uno muy abundante, el de las observaciones

típicas, y uno muy reducido, el de las discrepancias, y el desbalanceo, es decir, la diferencia de tamaño en ambos conjuntos es grande.

El esquema de detección de anomalías como un proceso de clasificación es útil en muchas circunstancias, pero es inadecuado para muchos otros casos. Esto se debe a que no es factible conocer con anticipación la clase o etiqueta de los datos, esto es, no se sabe si son o no anomalías. Este último caso es particularmente interesante. En última instancia, puede pensarse que detectarlas es un proceso equivalente a una clasificación de *una sola clase*, es decir, solamente contamos con las instancias de la clase usual. [9]

Las observaciones del fenómeno, estructura o proceso de interés suelen condensarse en una matriz de datos, en la que cada observación se representa como un renglón y las diferentes caracterizaciones suelen representarse como columnas; por lo que, la mayoría de los análisis computacionales se enfocan a escudriñar dicha estructura matricial. En cierto tipo de análisis, las observaciones se asocian a una clase externa, o etiqueta, lo que permite entrenar algún clasificador. Cabe mencionar que un clasificador es un algoritmo que, mediante la modificación de uno o más parámetros y con base en las caracterizaciones de los objetos, es capaz de decirnos si pertenece a una u otra clase.

Un objeto puede ser descrito de una o más formas, teniendo en cuenta que cada una de esas caracterizaciones es un atributo, rasgo o variable. Ahora bien, cada dato, instancia, punto o vector, indistintamente como los llamaremos, puede pensarse inmerso en un espacio de dimensionalidad igual al número de estas formas, es decir, cada caracterización define una dimensión. En la figura 1, cada punto es descrito por dos atributos: su posición a lo largo del eje horizontal y su ubicación a lo largo del eje vertical.

Las anomalías son detectadas a partir de la matriz de datos, también conocida por otros nombres como matriz de atributos, cubo de datos, conjunto de datos, muestra o base de datos. No obstante, lo relevante de estas es lo que contiene, ya que representa un conjunto de observaciones sobre algún fenómeno, proceso o estructura de interés.

En la gran mayoría de los casos, los datos que buscamos como potenciales discrepancias son multidimensionales. Por ejemplo, la tabla 1 muestra una lista de 40 ciudades en la República Mexicana, donde cada ciudad es descrita por 7 atributos: la precipitación pluvial en los meses de enero, junio y septiembre del 2020; las temperaturas ambiente

| Localidad | Lluvias Ene | Lluvias Jun | Lluvias Sept | Temp °C Ene | Temp °C Jun | Temp °C Sept | Altitud |
|----------------------|-------------|-------------|--------------|-------------|-------------|--------------|---------|
| Ecatepec | 4.50 | 91.10 | 84.73 | 15.02 | 19.32 | 18.30 | 2250 |
| León | 1.01 | 157.05 | 94.01 | 17.79 | 24.93 | 22.90 | 1815 |
| Puebla | 2.80 | 180.30 | 138.10 | 19.83 | 19.10 | 14.78 | 2135 |
| Guadalajara | 10.63 | 96.12 | 113.40 | 17.57 | 23.48 | 20.90 | 1566 |
| Monterrey | 36.32 | 38.51 | 310.92 | 15.34 | 30.79 | 28.70 | 540 |
| Chihuahua | 6.40 | 24.40 | 90.65 | 11.93 | 28.06 | 24.90 | 1415 |
| Mérida | 20.54 | 284.01 | 181.96 | 24.03 | 30.96 | 29.40 | 14 |
| Saltillo | 10.45 | 59.71 | 108.01 | 12.85 | 23.81 | 21.70 | 1560 |
| Hermosillo | 25.70 | 0.01 | 196.81 | 32.06 | 29.80 | 17.85 | 210 |
| SLP | 0.20 | 189.00 | 30.30 | 22.71 | 20.80 | 14.82 | 1864 |
| Culiacán | 0.51 | 2.41 | 301.25 | 29.66 | 28.10 | 19.24 | 280 |
| Querétaro | 0.00 | 229.40 | 96.20 | 23.42 | 21.90 | 16.36 | 1820 |
| Morelia | 19.82 | 178.03 | 55.81 | 16.02 | 21.44 | 20.00 | 1920 |
| Reynosa | 55.01 | 188.40 | 95.51 | 30.02 | 29.30 | 17.83 | 56 |
| Tlaquepaque | 11.83 | 124.23 | 112.45 | 18.73 | 26.66 | 24.50 | 1569 |
| Guadalupe | 38.20 | 97.80 | 403.80 | 14.22 | 29.75 | 27.90 | 500 |
| Durango | 1.21 | 8.14 | 96.31 | 14.57 | 25.83 | 22.20 | 1966 |
| Tlalpan | 5.59 | 207.27 | 76.20 | 15.87 | 15.65 | 15.50 | 2294 |
| Apodaca | 7.01 | 38.03 | 90.57 | 10.11 | 20.85 | 19.20 | 430 |
| Atizapán | 3.10 | 125.14 | 82.83 | 14.51 | 20.05 | 18.80 | 2348 |
| Matamoros | 34.52 | 145.51 | 106.32 | 30.30 | 29.50 | 19.59 | 39 |
| Gral. Escobedo | 31.25 | 12.75 | 356.00 | 15.48 | 30.85 | 28.70 | 510 |
| Xalapa | 27.01 | 261.53 | 268.62 | 15.72 | 22.37 | 21.40 | 1460 |
| Mazatlán | 0.00 | 0.42 | 116.62 | 28.47 | 29.90 | 20.79 | 13 |
| Miguel Hidalgo | 0.75 | 66.35 | 77.33 | 15.69 | 20.21 | 19.40 | 2273 |
| Sn Nicolas Garza | 36.32 | 38.51 | 310.92 | 15.97 | 30.56 | 28.90 | 512 |
| Veracruz | 31.92 | 180.80 | 239.24 | 23.06 | 28.95 | 28.80 | 10 |
| Celaya | 2.10 | 64.91 | 142.21 | 17.21 | 23.98 | 22.30 | 1767 |
| Tepic | 1.30 | 46.72 | 253.05 | 21.00 | 26.98 | 25.70 | 920 |
| Centro | 167.30 | 175.80 | 262.40 | 29.96 | 29.00 | 24.45 | 9 |
| Victoria | 26.80 | 70.60 | 86.50 | 16.61 | 30.00 | 28.30 | 318 |
| Cajeme | 14.51 | 0.01 | 87.03 | 31.57 | 30.20 | 18.77 | 44 |
| Soledad de G Sanchez | 0.01 | 185.50 | 28.60 | 22.20 | 20.20 | 13.50 | 1849 |
| Solidaridad | 99.62 | 24.20 | 61.70 | 32.55 | 31.20 | 27.30 | 10 |
| Sta. Catarina | 35.00 | 105.00 | 416.00 | 28.07 | 26.20 | 16.06 | 600 |
| Oaxaca | 0.40 | 112.40 | 102.42 | 24.78 | 23.80 | 20.66 | 1555 |

Tabla 1. 42 localidades de la República Mexicana descritas por siete atributos: precipitación pluvial total, temperaturas promedio en enero, junio y septiembre del 2020, todas en grados Celsius, y altitud promedio de la localidad.

promedio para los mismos tres meses y el séptimo atributo es la altitud promedio de la ciudad o alcaldía. Cada localidad es, por lo tanto, un punto en el espacio de dimensión 7. Como en la pantalla de la computadora únicamente podemos graficar en una, dos o tres dimensiones, y la mayoría

de los humanos podemos percibir de forma natural hasta tres dimensiones, es necesario recurrir a algún algoritmo de reducción de la dimensionalidad para darnos idea de cómo se distribuyen estos 40 puntos en el espacio de dimensión siete.

La figura 2 muestra una aproximación en dos dimensiones de la distribución de las ciudades en el espacio de atributos. El algoritmo utilizado, fue de escalamiento multidimensional. [12] El tamaño del nombre de la ciudad es indicador de qué tan anómalas es cada una de ellas, de acuerdo con lo que describiremos en breve. Mientras tanto, se invita al lector a observar con detenimiento las figuras 2 (a) y (b). Xalapa y Apodaca se muestran con tamaños de letra mayores en (a), lo que indica que su nivel de anomalía es mayor que el de otras ciudades. En (b), Apodaca disminuye su nivel, pero otras localidades lo aumentan, como es el caso de Tlalpan.

La función que evalúa el grado de anomalía, o el grado de cotidianidad de un objeto, se representa en general de manera implícita. Esta función se genera a partir de observaciones que constituyen el conjunto de entrenamiento. Como se mencionó anteriormente, la detección de anomalías puede verse como una clasificación de una sola clase, por ende, las observaciones anómalas son poco frecuentes o inexistentes. De esta forma, en el escenario más estricto, la función que clasifica a un objeto como usual o como discrepancia, se genera únicamente a partir de observaciones habituales. En la práctica, suponemos que el nivel de anomalía de un objeto es un continuo dentro de un

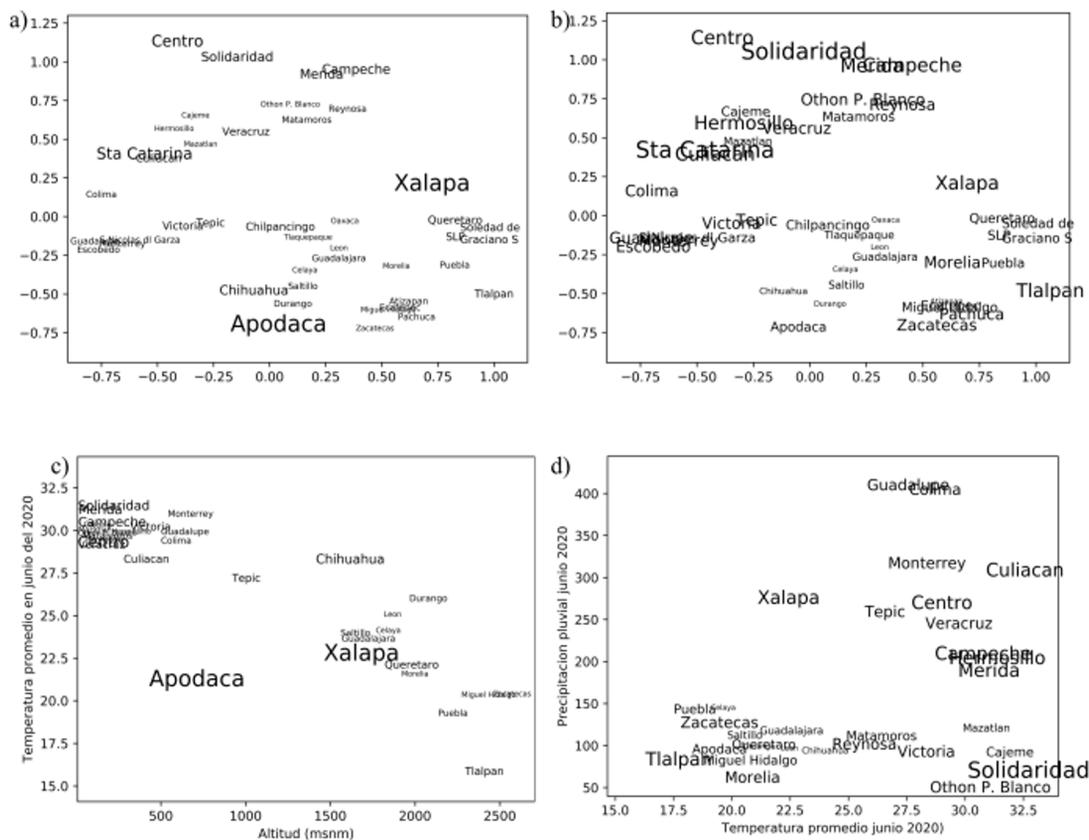


Figura 2. a) y b): Representación en dos dimensiones de las 42 ciudades de acuerdo con sus caracterizaciones en el espacio de siete atributos. La representación fue obtenida con escalamiento multidimensional, donde las ciudades que se encuentran próximas entre sí en la representación cuentan con descripciones semejantes en el espacio de atributos. Por ejemplo, Campeche y Mérida, mostradas como objetos muy semejantes entre sí, en realidad son descritas por atributos semejantes, como altas temperaturas en junio y septiembre, y no tan altas en enero, y con abundante precipitación en junio. Zacatecas y Pachuca son muy diferentes a las dos primeras, pero semejantes entre sí. El tamaño del nombre de la ciudad indica el nivel de anomalía de acuerdo con el algoritmo LOF. Se puede observar que Apodaca en Nuevo León, y Xalapa, en Veracruz, son mostrados como discrepancias. La razón de ello es que sus ciudades cercanas (en el espacio de atributos referido) son muy diferentes en cuanto a su vecindad a las dos ciudades. c) y d): Se muestra la proyección sobre dos de los siete atributos. Se observa que Apodaca tiene una temperatura más baja que las ciudades ubicadas en la misma altitud.

rango, usualmente entre 0 y 1. De esta forma, la dicotomía entre objeto usual y observación anómala desaparece, y se sustituye por un número real que proporciona a quien interpreta los datos, una mejor idea del nivel de anomalía de una observación.

De una matriz de datos, como la mostrada en la tabla 1 de nuestro ejemplo, es posible identificar posibles discrepancias bajo una de dos perspectivas. La primera evalúa cada punto con el resto, esto es, suponemos que serán identificadas de manera global. La segunda perspectiva compara a cada punto con los otros en su cercanía, lo que, en otras palabras, significa que serán detectadas de forma local.

El concepto de distancia es fundamental en muchos aspectos del aprendizaje computacional. No obstante, caracterizar el grado de anomalía de un objeto únicamente con el concepto de distancia puede llevar a caracterizaciones inestables. Para evitarlas, es necesario construir caracterizaciones más estables. Un concepto derivado del primero es la *densidad*. Ésta se define como el número de objetos que rodean a cada punto en el espacio de atributos. En la práctica, se define una vecindad en torno a cada punto, digamos, una esfera de radio r , y se lleva la cuenta, para cada punto, del número de objetos que se encuentran a una distancia r o menos.

A partir de la densidad, podemos darnos una idea de la distribución del número de vecinos de todos los objetos. El concepto anterior nos proporciona una buena aproximación a la diversidad de los datos, es mucho más poderoso, puede servir para comparar el nivel de semejanza entre vectores, y con ello, obtener un criterio de anomalía. A continuación, veamos cómo.

Algoritmo de Local Outlier Factor

Objetos cercanos en el espacio de atributos suelen mostrar la misma densidad. A Breunig y sus colaboradores, [13] se les ocurrió una idea interesante: se define la vecindad de un punto v como el conjunto de sus k objetos más próximos. Esta será indicada por la lista $L(k)$. A partir de $L(k)$, se obtiene alguna caracterización, por ejemplo, el cociente entre la distancia al vecino más alejado y al más próximo, dentro de los k puntos escogidos. Esta caracterización, digamos $C(k)$, servirá de base para la detección de anomalías locales. Ahora, se procede a caracterizar a los objetos en $L(k)$ de la misma forma. Al final, el nivel de anomalía del punto v es una función de qué tan diferente es su caracterización $C(k)$ de la de sus objetos cercanos. Si la caracterización de un objeto no se parece a la de sus objetos vecinos, ese objeto

puede constituir una anomalía. El algoritmo que Breunig y sus colaboradores propusieron lleva por nombre *Local Outlier Factor (LOF)*.

En la figura 2(a), el tamaño del nombre de la ciudad está dado por el nivel de anomalía computado por LOF. Cuanto más grande, mayor el valor de LOF, esto es, más anómala es la ciudad, en términos de la diferencia de densidad con sus $k = 3$ puntos más próximos. Las dos ciudades con mayor nivel de discrepancia son Apodaca en Nuevo León y Xalapa, en Veracruz. Ambas ciudades resultan anómalas, bajo los criterios de LOF, pues sus $k=3$ ciudades próximas, en el espacio de siete atributos, tienen una caracterización muy diferente. En términos coloquiales, Xalapa y Apodaca no se parecen a sus vecinas.

La amable comunidad lectora se preguntará por qué Xalapa y Apodaca son las localidades con mayor nivel de anomalía de acuerdo con LOF. La respuesta es que los vecinos de ambas ciudades en el espacio de atributos son muy diferentes a ellas en su caracterización. Sin embargo, podemos aproximar una respuesta intuitiva. En las figuras 2(c) y (d), se observa la gráfica de dispersión de las localidades estudiadas. Se muestra en (c) la altitud y la temperatura promedio en junio. En (d) se despliegan las localidades en el eje de temperatura promedio en junio y precipitación pluvial en el mismo mes. Para el primer caso, se observa que Apodaca se encuentra muy alejado de las demás localidades: ciudades con altitud semejante, la temperatura típica es realmente mayor. Mientras que, para el caso de Xalapa, en (d), ciudades con semejante temperatura promedio para el mes de junio, muestran una mínima precipitación pluvial.

Algoritmo de Bosques de Aislamiento (Isolation Forests)

Un segundo algoritmo, que parte de supuestos distintos, es el de Bosques de Aislamiento, o *Isolation Forests (IF)*. [14] Este algoritmo cuantifica el nivel de anomalía de un objeto en función de qué tan difícil es aislarlo del resto. Para ello, siempre en el espacio multidimensional de atributos, se trazará un hiperplano que sea perpendicular a alguno de los ejes. Los ejes o dimensiones que lo definen se eligen de manera aleatoria. Este hiperplano, puede ser visto como un aislamiento, y generará dos subregiones. Si el punto de interés es el único en la región, se habrá aislado de los demás; en caso contrario, el algoritmo se enfoca en la región que contiene al punto de interés y comenzará de nuevo el proceso de aislamiento. En el algoritmo IF, el número de iteraciones es el número de árboles o hiperplanos que son necesarios para aislar a un objeto del resto, es una medida

del nivel de anomalía. Cuantos más árboles sean necesarios, menos inusual es el objeto. Entre menos hiperplanos se requieran para aislar completamente a un objeto, mayor será su nivel de discrepancia.

Cabe mencionar que LOF es un algoritmo local de detección de anomalías, en tanto que IF es un algoritmo global. El segundo, compara el número de árboles o decisiones que fueron necesarias para aislar a un objeto de los demás, comparando esa cifra contra el de todos los demás. De esta forma, ordenan a los objetos con la dificultad de aislarlos del resto.

Es práctica común que al analizar datos unidimensionales se aplique una prueba estadística, esto con el fin de encontrar anomalías bajo el supuesto que los datos siguen una distribución gaussiana. La prueba de Grubbs es una técnica comúnmente utilizada, [15] en la que, si el criterio de rechazo de la prueba se cumple, se elimina el dato más extremo tomándolo como ruido o un error en la medición. Sin embargo, como se muestra en nuestro ejemplo de las localidades, no puede decirse que alguna de las ciudades de la figura sea ruido o un error en las mediciones, pero sí que represente una posible anomalía, si tomamos en cuenta los atributos que la describen. Un objeto detectado como anomalía al seguir un algoritmo dado, puede no ser considerado como tal al recurrir a otro algoritmo. No es que exista un error, se trata de satisfacer o no los supuestos detrás de cada algoritmo para suponer que un objeto es usual o no. En el caso de las localidades, Apodaca no es detectado como anomalía por *IF*, aunque sí lo es si nos basamos en los criterios detrás de *LOF*. Esto es enteramente válido no sólo en la detección de anomalías, sino en casi todas las técnicas de *análisis exploratorio de datos*.

Conclusiones

Una anomalía es una instancia, objeto, punto o vector que no guarda semejanza con el resto de las observaciones. Una anomalía no es sinónimo de ruido en los datos, o de errores en la medición. Es una observación potencialmente valiosa que puede dar luz a los especialistas sobre el fenómeno que se esté estudiando.

El proceso de detección de anomalías requiere de algún criterio cuantificable que sea semejante entre observaciones usuales y sustancialmente distinto en las instancias anómalas. Ese criterio es en general obtenido a partir de los atributos que describen a los objetos. Una manera de obtener ese atributo es por medio de algún algoritmo de

aprendizaje, en general no supervisado, pues la clase de las anomalías es o muy pequeña con respecto al tamaño de la clase usual o habitual, o inexistente.

Describimos dos algoritmos de detección de anomalías, basados en conceptos de densidad y distancia. Las anomalías pueden ser locales, o globales. Es local cuando difiere de sus vecinos cercanos, y lo es globalmente cuando es sustancialmente distinta a la mayoría de los vectores observados.

La inteligencia artificial busca, entre muchos otros objetivos, crear sistemas capaces de identificar, de manera no supervisada, clases o categorías de objetos semejantes bajo algún criterio latente. Una anomalía es, por definición, una instancia distinta a las observaciones usuales, y la búsqueda de algoritmos capaces de identificarlas es una tarea abierta.

Los autores de este trabajo forman parte del grupo Anomalocaris, dedicado al desarrollo de algoritmos de detección de anomalías, y su aplicación en diferentes contextos. La liga del sitio se encuentra en: <https://github.com/antonioneme/anomalocaris>.

Agradecimientos

N.H. y A.N. agradecen a la Dirección General de Asuntos de Personal Académico (DGAPA) por el apoyo a sus proyectos de investigación PAPIIT, con número IT100220 y IA103921, respectivamente.

BIBLIOGRAFÍA

- [1] M.A.F. Pimentel, D.A. Clifton, L. Clifton, *et al.*, “A review of novelty detection,” en *Signal processing*, vol.99, pp. 227-236, Jun. 2014. <https://doi.org/10.1016/j.sigpro.2013.12.026>
- [2] M. Goldstein, and S. Uchida, “A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data,” en *PLoS ONE*, vol.11, no. 4: E0152173, Apr. 2016. <https://doi.org/10.1371/journal.pone.0152173>
- [3] X. Xu, H. Liu, and M. Yao, “Recent Progress of Anomaly Detection,” en *Complexity*, vol.2019, Article ID 2686378, p.11, Jan. 2019. <https://doi.org/10.1155/2019/2686378>
- [4] H. Zhao, *et al.*, “Anomaly Detection for Medical Images using Self-supervised and Translation-consistent Features,” en *IEEE Transactions on Medical Imaging*, 2021. [10.1109/TMI.2021.3093883](https://doi.org/10.1109/TMI.2021.3093883)
- [5] L. Selicato, *et al.*, “Ensemble Method for Detecting Anomalies in Gene Expression Matrices,” en *Mathematics* 2021, vol. 9, p. 882. <https://doi.org/10.3390/math9080882>
- [6] A. Neme, B. Lugo y A. Cervera “*Authorship attribution as a case of anomaly detection: A neural network model*,” en *Int. J. Hybrid Intell. Syst.* Vol.8, no. 4, pp. 225-235, 2011.
- [7] P. Garrard, L. Maloney, J. Hodges, *et al.*, “The effects of very early Alzheimer’s disease on the characteristics of writing by a renowned author,” en *Brain*. 2005 Feb;128(Pt 2):250-60. [10.1093/brain/awh341](https://doi.org/10.1093/brain/awh341)
- [8] M. Markou y S. Singh. “Novelty detection: a review—part 1: statistical approaches,” vol.83, no. 12, pp. 2481-2497, 2003. <https://doi.org/10.1016/j.sigpro.2003.07.018>
- [9] M. Markou y S. Singh. “Novelty detection: a review—part 2: neural network based approaches,” vol. 83 no.12, pp. 2499-2521, 2003. <https://doi.org/10.1016/j.sigpro.2003.07.019>
- [10] N. Nilsson. *The Quest for Artificial Intelligence 1st Edition*. Cambridge University Press, 2009. ISBN-13: 978-0521122931.
- [11] G. James, D. Witten, T. Hastie, *et al.*, “An introduction to statistical learning with applications in R, second edition,” en *Springer*, 2021.
- [12] JB. Kruskal, “Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis,” en *Psychometrika*, vol. 29, no.1, pp. 1–27, 1964. doi:10.1007/BF02289565

- [13] M. Breunig, H. Kriegel, R. T. Ng, *et al.*, “LOF: identifying density-based local outliers,” en *ACM SIGMOD Record*, vol.29, no. 2, pp. 93–104, June 2000. <https://doi.org/10.1145/335191.335388>
- [14] T. Liu, F. Tony, T. Kai Ming, *et al.*, “Isolation Forest,” en *2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422. doi:10.1109/ICDM.2008.17. ISBN 978-0-7695-3502-9. S2CID 6505449, 2008.
- [15] F. E. Grubbs, “Sample criteria for testing outlying observations,” en *Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 27–58. doi:10.1214/aoms/, 1950

Cómo se cita:

N. Hevia, S. Mota y A. Neme, “Las anomalías: ¿qué son?, ¿dónde surgen?, ¿cómo detectarlas?” *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 4, octubre, 2021. [En línea]. Disponible en: <http://ties.unam.mx/> [Consultado en mes día, año].

Fecha de recepción: agosto 2, 2021

Fecha de publicación: octubre de 2021

APRENDIZAJE COMPUTACIONAL PARA ANÁLISIS DE IMÁGENES DE ULTRASONIDO MÉDICO

F. Arámbula Cosío, F. Torres Robles, G.A.R. Velásquez-Rodríguez, E. Galicia Gómez, B. Escalante-Ramirez, J. Olveres, Jorge L. Pérez, V. Medina Bañuelos, L. Camargo Marín y M. Guzmán Huerta

<http://www.ties.unam.mx/>

Fecha de recepción: agosto 10, 2021 • Fecha de publicación: octubre de 2021

Octubre 2021 | número de revista 4 • ISSN en trámite



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)

APRENDIZAJE COMPUTACIONAL PARA ANÁLISIS DE IMÁGENES DE ULTRASONIDO MÉDICO

Resumen

Las imágenes de ultrasonido médico ofrecen varias ventajas con respecto a otras modalidades como rayos X, tomografía computarizada y resonancia magnética nuclear. Los equipos de ultrasonido son compactos y el tamaño pequeño de la sonda permite al experto tomar imágenes de una amplia variedad de órganos en 2 y 3 dimensiones. Las principales desventajas son un alto nivel de ruido (*speckle*) en las imágenes y la necesidad de que un experto realice la adquisición de imágenes. En este trabajo presentamos cinco sistemas de análisis computarizado de imágenes de ultrasonido, basados en técnicas de aprendizaje computacional, para mejorar la calidad de la detección de estructuras de interés clínico. Reportamos aplicaciones en: ecocardiografía, salud materno-fetal y oncología.

Palabras clave:

segmentación del ventrículo izquierdo, segmentación del cerebelo fetal, composición espacial de ultrasonido 3D, detección de huesos en ultrasonido.

COMPUTERIZED MEDICAL ULTRASOUND IMAGE ANALYSIS

Abstract

Medical ultrasound has several advantages with respect to other imaging modalities such as x rays, computed tomography and magnetic resonance imaging. Ultrasound scanners are compact and the small size of the probe allows for the scanning of a wide variety of anatomical structures in two and three dimensions. Its main limitations are high levels of speckle noise and the need for an expert scanning of the patient. In this work we report five systems of computerized medical image analysis, based on machine learning techniques, to improve the detection of anatomical structures of clinical interest

Keywords:

segmentation of the left ventricle, segmentation of the fetal cerebellum, spatial compounding of 3D ultrasound, bone detection in ultrasound.

APRENDIZAJE COMPUTACIONAL PARA ANÁLISIS DE IMÁGENES DE ULTRASONIDO MÉDICO

I. Introducción

En este artículo presentamos varios sistemas, basados en métodos de aprendizaje computacional, con métodos estadísticos, para el análisis de imágenes de ultrasonido (US) médico en 2 y 3 dimensiones, con aplicaciones en: ecocardiografía, salud materno-fetal y oncología. Todos los sistemas son resultados de la colaboración de dos instituciones de educación superior (Universidad Nacional Autónoma de México y Universidad Autónoma Metropolitana) y una institución de salud pública (Instituto Nacional de Perinatología), en México.

El ultrasonido médico ofrece varias ventajas con respecto a otras modalidades de imágenes médicas (por ejemplo, rayos X, tomografía computarizada y resonancia magnética). Los equipos tienen dimensiones reducidas y la adquisición de imágenes es muy versátil. Un experto puede explorar diferentes regiones anatómicas con facilidad. No se utiliza radiación ionizante para formar la imagen – como es el caso con rayos X-, por lo que es posible adquirir imágenes de ultrasonido de cualquier paciente con toda seguridad. Los costos de los equipos son los más bajos de todas las modalidades de imágenes médicas disponibles. Las principales desventajas de las imágenes médicas de ultrasonido, son el alto contenido de ruido de moteado (*speckle*) [1] y la necesidad de que un experto médico realice simultáneamente la adquisición de imágenes y la interpretación en el paciente (es decir, es una modalidad de imágenes médicas que depende del usuario).

La penetración del ultrasonido en el cuerpo humano, para la formación de imágenes, está limitada a unos 15 cm, que permiten todavía visualizar una buena parte de la anatomía humana. Los avances continuos en la adquisición, el procesamiento de señales y la formación de imágenes, han aumentado constantemente la calidad, reduciendo los niveles de ruido *speckle*. [2] Actualmente, el ultrasonido es la modalidad más utilizada en medicina, [3] y nos permite, en algunos casos, visualizar estructuras de tejido blando que no se pueden ver con rayos X. Por ejemplo, los tumores de mama en mujeres jóvenes pueden muchas veces detectarse mejor con ultrasonido que con rayos X (mastografía). Los equipos modernos permiten visualizar imágenes (2D) o volúmenes (3D), estáticos o como secuencias de video, de la anatomía del paciente.

Por sus altos niveles de seguridad y accesibilidad, el ultrasonido se ha utilizado activamente en el monitoreo de la salud materno fetal durante la gestación. También se utiliza mucho en ecocardiografía (fetal y posnatal en todas las edades), hígado, próstata, músculos, vascularidades y recientemente para detectar la superficie de algún hueso de interés durante cirugías ortopédicas. En este artículo presentamos varios sistemas de análisis de imágenes de ultrasonido médico, desarrollados en colaboraciones entre la Facultad de Ingeniería y el IIMAS de la UNAM, la UAM Unidad Iztapalapa y el Instituto Nacional de Perinatología (INPER) en la Ciudad de México. Los sistemas se basan en técnicas de aprendizaje computacio-

nal (*machine learning*), para realizar automáticamente alguna tarea de apoyo al diagnóstico médico en cardiología, perinatología (salud materno-fetal) y oncología de mama, como se describe a continuación.

II. Anotación automática del ventrículo izquierdo

Uno de los índices más importantes del buen funcionamiento del corazón es la fracción de eyección, calculada como el volumen de sangre que expulsa el ventrículo en cada contracción. La fracción de eyección se mide utilizando imágenes de ecocardiografía (ultrasonido) del ventrículo izquierdo (figura 1). Para una medición exacta, es necesario anotar el contorno del ventrículo cuando tiene el máximo volumen de sangre (diástole) y cuando tiene el volumen más pequeño (sístole).

En un trabajo de colaboración entre el Laboratorio Avanzado de Procesamiento de Imágenes de la Facultad de Ingeniería y la Unidad Yucatán del IIMAS de la UNAM, desarrollamos un sistema automático para la anotación del ventrículo izquierdo en imágenes de ultrasonido, que nos permitirá el cálculo automatizado de la fracción de eyección. A continuación, describimos el método de modelado estadístico, de las formas que puede tomar un órgano, que utilizamos para la anotación automática del ventrículo izquierdo.

II.1. Modelo de distribución de puntos y modelo de formas activas (*active shape model*) del ventrículo izquierdo

Un modelo de distribución de puntos (MDP) [4] es una manera efectiva de modelar matemáticamente la variedad de formas que puede tomar el contorno de uno o varios órganos en 2D (imágenes) y en 3D (volúmenes). El modelo se construye a partir de un conjunto de ejemplos de las formas de interés. Sobre el contorno o la superficie de cada ejemplo, se muestrea un número (n) de puntos de referencia (*landmarks*), como se ilustra en la figura 2, con ejemplos de formas del ventrículo izquierdo.

Las coordenadas (x,y) de todos los puntos de referencia de cada ejemplo de entrenamiento, se almacenan en un vector. Todos los ejemplos se alinean dentro del mismo sistema de coordenadas de referencia, como se describe en la referencia 4. Con todos los vectores alineados del conjunto de entrenamiento, podemos fácilmente calcular la forma promedio como el vector promedio, como se ilustra en la figura 2.

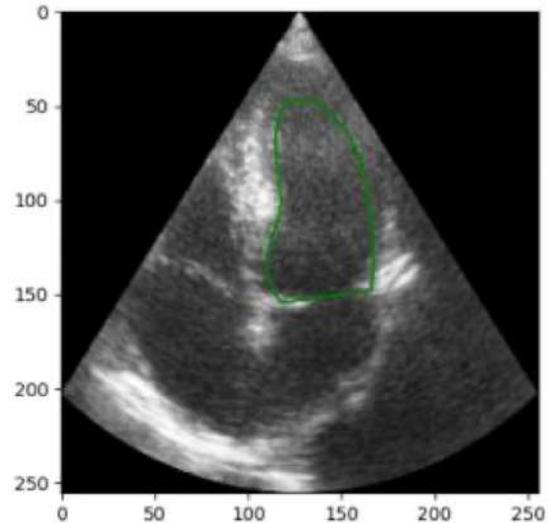


Figura 1. Ultrasonido del ventrículo izquierdo con la anotación del experto

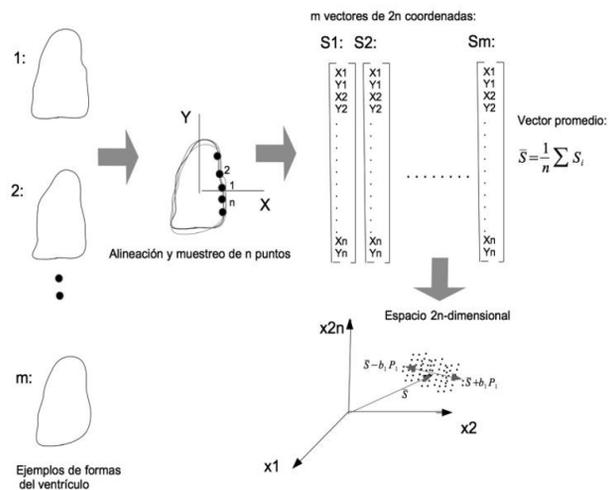


Figura 2. Construcción de un modelo de distribución de puntos (MDP) del ventrículo izquierdo

Así como la representación gráfica de un vector de tres coordenadas (x,y,z) corresponde a un punto en un sistema de 3 dimensiones (x,y,z), la misma interpretación se puede aplicar a vectores con cualquier número de coordenadas: un vector de n coordenadas corresponde a un punto en un espacio de n dimensiones. Nuestra forma promedio calculada anteriormente, corresponde a un punto en el espacio

de $2n$ -dimensiones (n puntos con dos coordenadas x,y), como se ilustra en la figura 2. Como todos los vectores de nuestros ejemplos del ventrículo son similares, esperamos que los puntos correspondientes ocupen una región compacta en el espacio de $2n$ -dimensiones. Podemos entonces modelar la forma aproximada y el tamaño de esta región, con un análisis de componentes principales (ACP) de todos los vectores, como se describe en la referencia 4, y podemos generar nuevas formas del ventrículo como la suma del vector (ventrículo) promedio \bar{s} más un incremento b alrededor del promedio, dentro de la región de “ventrículos” que determinamos con el ACP, como se ilustra en la figura 2 para un incremento $b_1 P_1$, donde b_1 es un escalar y P_1 es un vector (abajo a la derecha en la figura 2). Las formas correspondientes se muestran junto con la forma promedio del ventrículo en la figura 3.

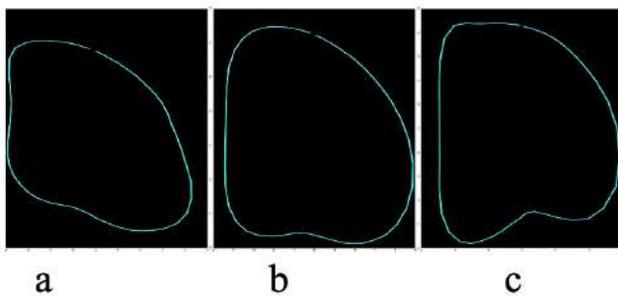


Figura 3. Formas del MDP del ventrículo izquierdo para: a) $s = \bar{s} - b_1 P_1$; b) $s = \bar{s}$; c) $s = \bar{s} + b_1 P_1$

Cootes y Taylor [4] desarrollaron también un método de ajuste automático del MDP a una nueva imagen, que bautizaron como modelo de formas activas (*active shape model*). El método de ajuste se basa en la búsqueda de bordes cercanos a la forma inicial del MDP. Con la detección de bordes se ajusta la forma del MDP y se repite la búsqueda y el ajuste de un número fijo de iteraciones. Usualmente, basta con 50 iteraciones para lograr un buen ajuste. En la figura 4 mostramos 3 ejemplos de resultados del ajuste automático del MDP del ventrículo izquierdo, entrenado con 50 ejemplos.

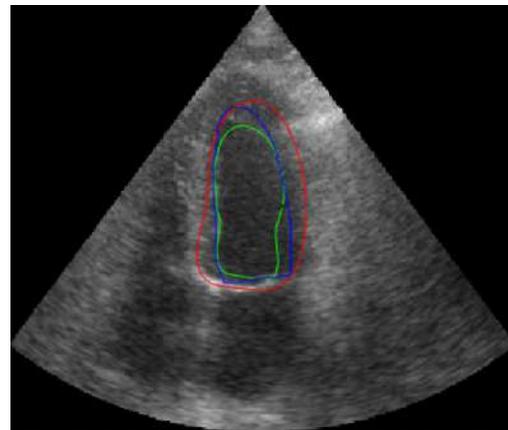
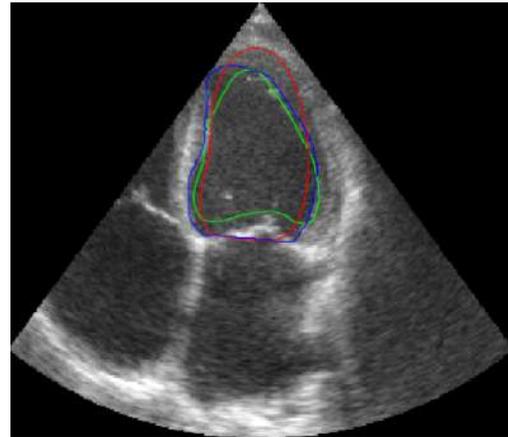


Figura 4. Resultados de la anotación automática del ventrículo izquierdo, en tres imágenes de validación: rojo, inicialización de la forma promedio; azul, la anotación del experto, y verde, la segmentación automática final.

III. Ultrasonido del Cerebro Fetal

Como hemos mencionado, el ultrasonido es una técnica no invasiva, que trabaja en tiempo real, no requiere el uso de radiación ionizante, el equipo es compacto y portátil, el transductor es pequeño y fácilmente manipulable. Estas cualidades hacen que sea la modalidad más utilizada en obstetricia, para la valoración de la salud materno-fetal. En particular, el estudio del cerebro fetal mediante imágenes de ultrasonido (US), puede ayudar a los expertos obstetras a dar un seguimiento del neurodesarrollo del feto. Sin embargo, la ecografía presenta algunas limitaciones que pueden restringir la adecuada medición de diversos índices y, como consecuencia, dificultar un buen diagnóstico y valoración de la salud fetal. El ultrasonido presenta variaciones entre pacientes y entre equipos; las imágenes no tienen contornos bien definidos y muestran ruido multiplicativo (*speckle*) con variaciones en intensidad, y puede presentar diversos artefactos y sombras acústicas. [5] A continuación, presentamos tres sistemas de análisis de imágenes de US del cerebro fetal –basados en métodos de aprendizaje computacional–, que contribuyen al diagnóstico clínico exacto y eficiente.

III.1. Medición automática del pliegue nucal

Durante la gestación, el pliegue nucal se forma por la acumulación normal de líquido en la base del cráneo del feto, sin embargo, un grosor fuera de rangos normales del pliegue nucal es el principal indicador para la detección de síndrome de Down, durante el segundo trimestre de la gestación. En esta sección presentamos la construcción de un modelo de formas activas (*active shape model*), para la medición automática del grosor del pliegue nucal, en imágenes de ultrasonido del cerebro fetal.

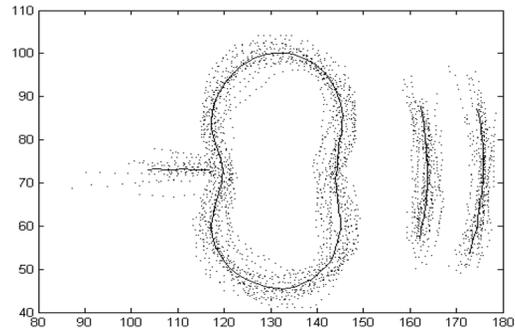


Figura 5. Forma promedio de las estructuras del pliegue nucal (línea continua) y los puntos muestreados de 20 ejemplos de entrenamiento.

III.1.1. Modelo de distribución de puntos y modelo de formas activas (*active shape model*) del pliegue nucal

Un modelo de distribución de puntos puede también representar las formas de varios objetos. En la figura 5 mostramos los ejemplos de entrenamiento y la forma promedio de todas las estructuras anatómicas de interés durante la medición del pliegue nucal –línea media, cerebelo, placa occipital y borde de la piel–, con los puntos muestreados correspondientes, que permitirán la construcción del MDP del pliegue nucal. Los vectores de entrenamiento se formaron con las coordenadas (x,y) de todos los puntos de las 4 estructuras anatómicas que se ilustran en la figura 5.

En la figura 6 mostramos cómo se deforma el MDP de la “plantilla del pliegue nucal”.

En la figura 7 mostramos dos ejemplos del resultado del ajuste automático del MDP del pliegue nucal, a partir de una forma inicial.

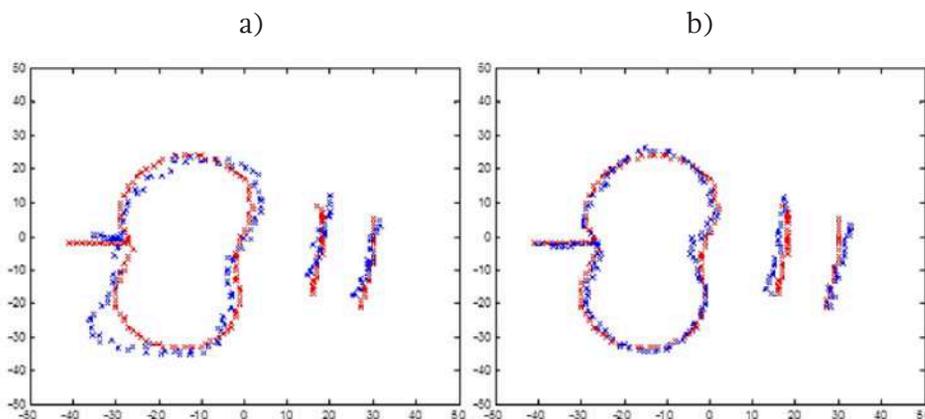


Figura 6. Formas del MDP del pliegue nucal para: a) $s=s-b_1 P_1$; b) $s=s+b_1 P_1$; se muestra en rojo la forma promedio y en azul la nueva forma.

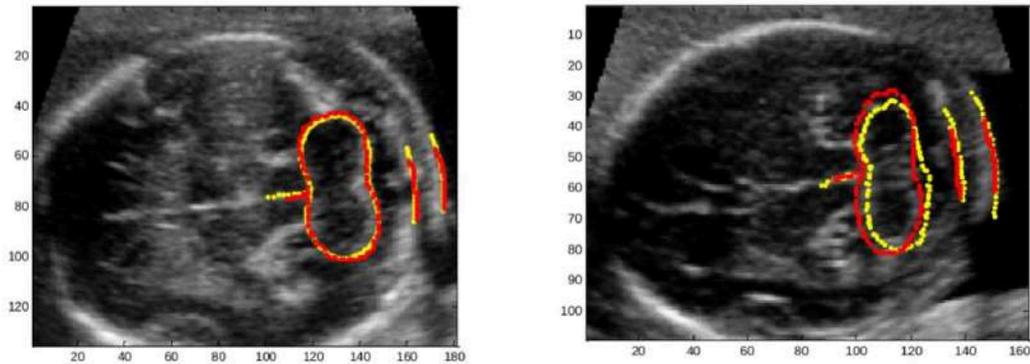


Figura 7. Resultados del ajuste automático del ASM del pliegue nucal: rojo, anotación experta; amarillo, ASM

III.2. Anotación automática del cerebelo en 3 dimensiones (3D)

El cerebelo es una estructura cerebral importante, por su participación en las funciones cognitivas, como la atención y el lenguaje, así como la coordinación de movimientos, precisión y sincronización precisa. [6] También se puede utilizar para evaluar la edad gestacional o afecciones prenatales, como las anomalías en la fosa posterior craneal y defectos del tubo neural. [6,7] La segmentación automática de estructuras cerebrales en la modalidad de ultrasonido, es una tarea desafiante debido a los ar-

tefactos, como el “speckle” y la poca homogeneidad de intensidad, debido a los movimientos fetales durante los procesos de adquisición de imágenes.

Para la segmentación automática del cerebelo en 3D, una experta en perinatología realizó la anotación manual de los 15 principales cortes que conforman el cerebelo (planos axiales), en 20 volúmenes de ultrasonido, con el fin de generar posteriormente un modelo de distribución de puntos de la forma del cerebelo en tres dimensiones (3D) y sus modos de variación, que nos permiten un buen

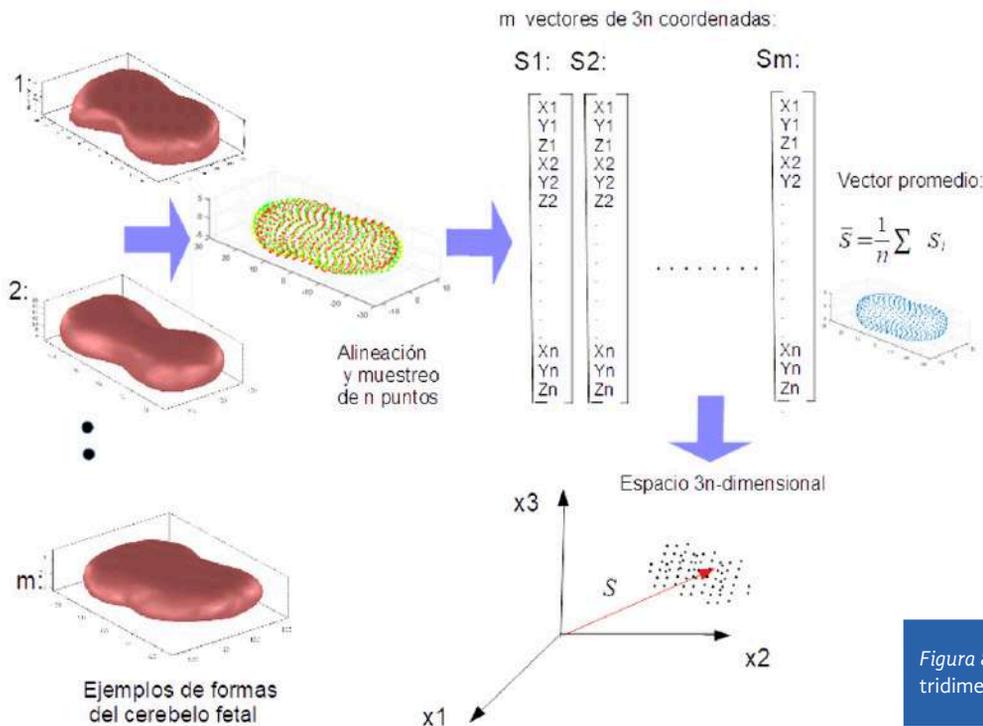


Figura 8. Construcción del MDP tridimensional del cerebelo fetal

ajuste en un nuevo volumen. [8] Un MDP de formas en tres dimensiones se construye de manera análoga a un modelo de formas en 2 dimensiones, como lo presentamos en la figura 1. Sin embargo, ahora se muestrean puntos sobre la superficie de cada ejemplo de forma 3D y se almacenan las coordenadas (x,y,z) de todos los puntos en un vector. La forma promedio y los modos de variación de forma, se calculan con el mismo análisis de componentes principales, descrito anteriormente para el MDP del ventrículo izquierdo (figura 8).

El ajuste del MDP 3D se basó en una función objetivo, que depende de los parámetros de forma y postura (ángulo, escala y traslaciones) del MDP, para la detección de bordes que corresponden al cerebelo, dentro de un volumen de ultrasonido del cerebro fetal. En nuestros resultados observamos que la anotación automática del cerebelo -utilizando el MDP 3D- produce en promedio un 90% de correspondencia con los volúmenes anotados por una experta, como se ilustra en la figura 9, donde se muestra en amarillo la anotación tridimensional de la experta y en azul la anotación automática.

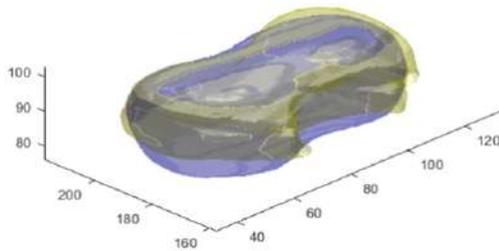


Figura 9. Comparativa de la segmentación de la experta (amarillo) contra la automática (azul)

III.3. Composición espacial de imágenes cerebrales fetales de US 3D

En el caso particular de la neuroecografía fetal, existe un problema de oclusión causada por la calcificación y el endurecimiento craneal durante el segundo y el tercer trimestre gestacional, lo que impide el paso del sonido al interior del cráneo. Esto repercute en la aparición de sombras en las regiones correspondientes al cerebro fetal. [9,10] Este fenómeno impide la correcta valoración del cerebro, así como la medición de diversas estructuras internas. En la figura 10 se muestran dos ejemplos de estudios cerebrales fetales con sombras acústicas. Se puede notar claramente la ausencia de información correspondiente al tejido cerebral.

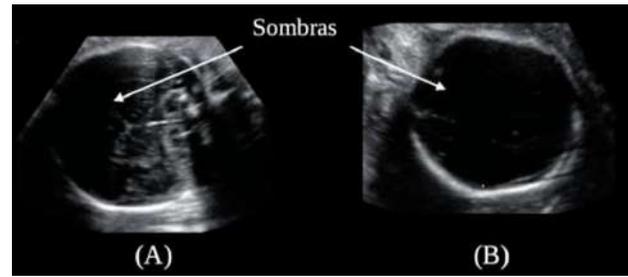


Figura 10. Ejemplos de imágenes de US fetal cerebral con artefactos de oclusión acústica

Para lidiar con el problema de las oclusiones acústicas cerebrales, se ha propuesto una estrategia basada en algoritmos de procesamiento digital de imágenes y aprendizaje computacional. En la figura 11 se muestra la metodología propuesta, con el objetivo de mejorar la calidad de las imágenes cerebrales fetales de US en 3D, disminuyendo la cantidad de sombras acústicas.

La metodología comienza con la adquisición de dos o más estudios de US 3D de la cabeza fetal. El propósito consiste en hacer la composición o fusión de varios estudios de US, para generar uno nuevo de mejor calidad y menor cantidad de sombras acústicas. De acuerdo con lo reportado por Timor-Trisch *et al.*, [9] diferentes ángulos de adquisición pueden brindar información complementaria del tejido cerebral. Por lo tanto, la metodología comienza con la adquisición de estudios cerebrales desde dos ángulos: axial y coronal, como se puede observar en la figura 11A. El siguiente paso es alinear ambos volúmenes en un mismo plano de referencia, para lo que se propuso un algoritmo de alineación o registro automático basado en bosques aleatorios (figura 11.B). [11] Para hacer una adecuada composición o fusión de los estudios cerebrales, se propuso estimar un factor de ponderación. Este factor tiene como objetivo asignar valores cercanos a "1" a las regiones de la imagen que contengan información del tejido cerebral y valores cercanos a "0" a las regiones que contengan sombras acústicas. Para lograr esto, se propone estimar mapas probabilísticos para cada volumen de US a fusionar. Estos mapas probabilísticos fueron obtenidos mediante una máquina de soporte vectorial (MSV), alimentada con un conjunto de 264 mapas de textura llamados matrices de coocurrencia de escala de grises (MCEG), los cuales fueron seleccionados con un análisis de componentes principales (ACP) (figuras 11C y 11D). La MSV es un algoritmo de aprendizaje computacional, el cual fue

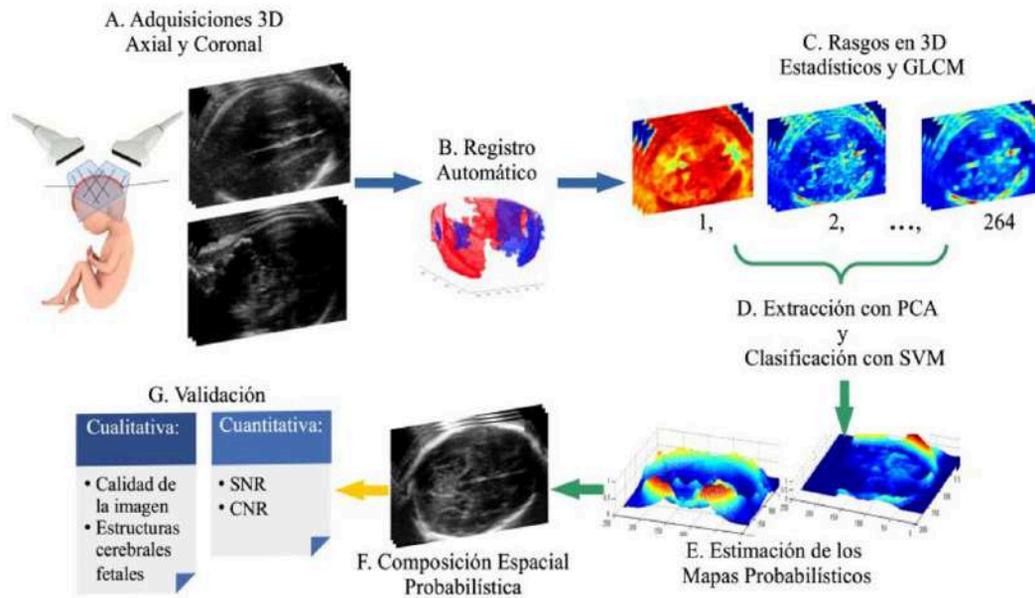


Figura 11. Diagrama general de la metodología propuesta para la composición de estudios cerebrales fetales de US 3D [7]

entrenado para clasificar y ponderar regiones con tejido cerebral y zonas con oclusiones acústicas, generando así los mapas de ponderación (figura 11.E). Para hacer la fusión de los diversos volúmenes, se utilizó la norma euclidiana ponderada con los mapas previamente obtenidos. Como resultado, se obtiene un nuevo estudio cerebral fetal de US 3D, que contiene la información de los estudios fusionados y ponderados de forma probabilística (figura 11.F).

Finalmente, para evaluar el desempeño, medimos la relación señal a ruido (SNR, *signal to noise ratio*) y la relación contraste a ruido (CNR, *contrast to noise ratio*), las cuales mejoraron de $7.8+/-2.8$ a $10.5+/-2.7$ para SNR y de $5.1+/-1.9$ a $12.1+/-2.3$ para CNR. Esto refleja cuantitativamente la mejoría en la calidad del nuevo estudio generado. Además, expertos en obstetricia efectuaron una evaluación cualitativa, quienes mencionan que estructuras como la línea media (LM), cerebelo (CER) o circunferen-

cia craneal (CC), se pueden apreciar con mejor calidad en el estudio de US, resultado de la composición. En la figura 12 se muestra la comparación entre un estudio tradicional y el resultado del método propuesto. En la figura 12.A se observan algunas sombras acústicas. En contraste, en la figura 12.B se puede observar una mayor cantidad de tejido cerebral, así como una mejor calidad en diversas estructuras.

Se considera que el algoritmo propuesto puede ayudar a generar un nuevo estudio cerebral fetal de US 3D de mejor calidad, que permita a los expertos en obstetricia efectuar una mejor valoración del cerebro, así como facilitar las mediciones de diversas estructuras, como el cerebelo o la circunferencia craneal. Como trabajo a futuro, se pretende seguir explorando nuevos algoritmos basados en redes neuronales artificiales, de los cuales ya se tienen resultados preliminares. [12]

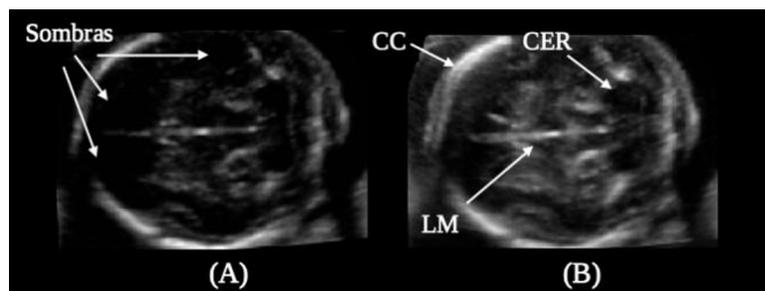


Figura 12. Ejemplo de una imagen de US fetal cerebral sin composición (A) y con el algoritmo de composición propuesto (B). CC: circunferencia craneal, CER: cerebelo y LM: línea media

IV. Anotación automática de tumores de mama en ultrasonido 2D

El cáncer de mama se ha convertido en la causa número uno de muertes en mujeres alrededor del mundo. La detección temprana y la estimación de riesgo son críticas para mejorar el pronóstico del paciente. Aunque la resonancia magnética es la herramienta recomendada por la OMS para la visualización del cáncer de mama, el ultrasonido se ha convertido en una herramienta muy útil para asistir y complementar este procedimiento, principalmente en mujeres jóvenes, en las cuales los tumores tienden a ser más grandes y malignos, ya que es capaz de detectar las lesiones en tejido mamario denso, usualmente presente en mujeres menores a 30 años, con mayor sensibilidad que la mamografía. Otra ventaja es que es la mejor herramienta para distinguir entre masas sólidas, como fibroadenomas benignos y cáncer, y quistes benignos llenos de líquido, dando indicios de la malignidad de la masa. El ultrasonido puede ser utilizado también para guiar procedimientos quirúrgicos, como las biopsias de tumores de mama, procedimiento que es más exacto y confiable para el diagnóstico de cáncer. [13]

Se han realizado diversos esfuerzos para efectuar la segmentación automática de lesiones en imágenes de ultrasonido, con la finalidad de localizar y delinear los contornos, y así ayudar a los médicos en la visualización, la detección y el diagnóstico de cáncer. Dentro de los métodos propuestos destacan los basados en el aprendizaje computacional. Estos métodos tienden a ser más robustos, ya que toman en cuenta más de una característica para la identificar lesiones. Métodos supervisados como redes Neuronales Artificiales, *AdaBoost* y *Random Forest*, son comúnmente utilizados para la clasificación de píxeles y así poder realizar la diferenciación del tejido lesionado y el tejido adyacente. [14, 15] Sin embargo, las características extraídas de la imagen y los parámetros de entrenamiento deben ser cuidadosamente elegidos y optimizados para obtener el mejor modelo de aprendizaje posible. El patrón interno de eco tiene mayor poder descriptivo que el nivel de intensidad de gris. Diferentes descriptores de textura de primer orden (momentos centrales del histograma) [16], segundo orden (descriptores basados en matrices de coocurrencia y *run-length*) [17] y otros (transformada de Fourier y Hermite), [18] han sido utilizados

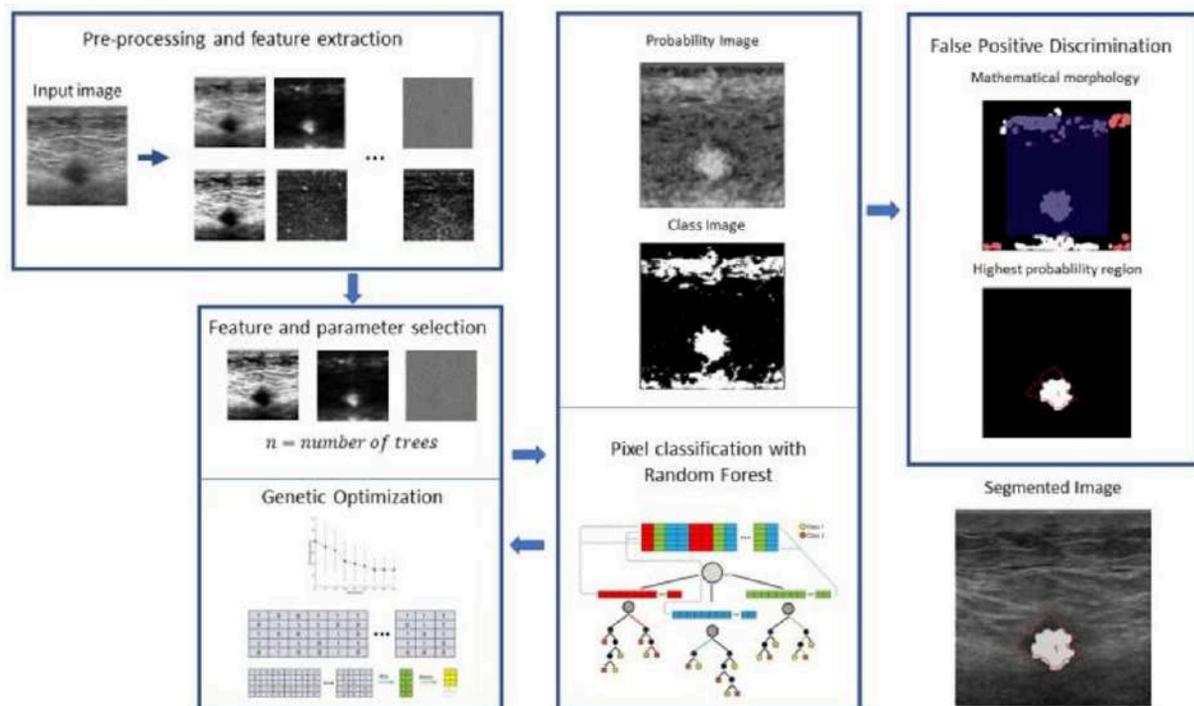


Figura 13. Metodología de segmentación de lesiones de mama en imágenes de ultrasonidos

para caracterizar las texturas de las lesiones de mama en imágenes de ultrasonido. En la correcta selección de características y la reducción de dimensiones para la optimización de los resultados en la clasificación, se han propuesto diversos métodos como el Análisis de Componentes Principales (ACP) y la optimización genética. [19]

Aunque los métodos de aprendizaje computacional son robustos, usualmente no pueden producir segmentaciones exactas y el refinamiento de la segmentación, como la eliminación de regiones falso-positivas, es necesaria. [20] Imágenes de probabilidad, creadas a partir de los resultados en la estimación de probabilidad de clase, obtenida mediante los métodos de aprendizaje, pueden ser utilizadas para la eliminación de regiones falso-positivas y desconectar regiones débilmente conectadas. En la figura 13 se presenta una metodología completa para la segmentación de lesiones de mama en imágenes de ultrasonido, incluyendo la selección de características y parámetros óptimos, para entrenar modelos de aprendizaje computacional, optimizado con un algoritmo genético simple y el posprocesamiento necesario para la eliminación de regiones falso-positivas. Todas las etapas del método se entrenaron con la validación cruzada en 88 imágenes con $k=4$ (25% de imágenes de validación).

Esta metodología incluye la extracción de características de textura de primer y segundo orden, así como la selección de características, la reducción de dimensión y la selección de parámetros óptimos. Posteriormente, se realiza la eliminación de regiones falso-positivas, mediante el uso de la imagen de probabilidad.

En la tabla 1 se presenta el modelo con los mejores resultados de exactitud, sensibilidad y especificidad, y los parámetros y descriptores de textura seleccionados por el algoritmo genético para la optimización de los resultados.

V. Comentarios finales

En este artículo presentamos 3 sistemas, basados en técnicas de aprendizaje computacional, para la anotación automática de estructuras anatómicas, como: el ventrículo izquierdo, el pliegue nucal y el cerebelo. Presentamos la aplicación de un método de modelado estadístico de formas: *active shape model* en 2D –ventrículo y pliegue nucal- y en 3D -cerebelo-. El método modela (o aprende) la variedad de formas que pueden tomar las estructuras anatómicas, a partir de un conjunto de ejemplos. Presentamos también un sistema de mejora de la calidad de las imágenes de ultrasonido del cerebro fetal, que fusiona varias vistas de la misma región con un método de aprendizaje probabilístico, logrando así una mejor visualización de las estructuras cerebrales sin las sombras que produce el cráneo del feto en desarrollo. Finalmente, presentamos también un sistema para la anotación automática de tumores de mama, basado en la extracción de características de textura de primer y segundo orden, así como la selección de características, la reducción de dimensión y la selección de parámetros óptimos. A través del cálculo de una imagen de probabilidad, se estima con buena exactitud la región que ocupa el tumor en la imagen de ultrasonido.

| Clasificador | Exactitud | Sensibilidad | Especificidad | Parámetros | Descriptores |
|----------------------|-----------|--------------|---------------|---|---|
| <i>Random Forest</i> | 82.92 | 82.23% | 82.61% | Arboles = 19 850 m = 10 max_sampl e = 1 | 19 descriptores 2 histograma 3 coocurrencia 4 run-length 10 hermite |

Tabla 1. Resultados de la segmentación mediante optimización genética

Todos los sistemas reportados en este artículo están concebidos para apoyar a los especialistas en el diagnóstico y la toma de decisiones clínicas. El tipo de métodos basados en modelos estadísticos de formas y texturas permite la predicción, con buena exactitud, a partir de algunas decenas de ejemplos de entrenamiento. Esto a su vez hace viable la mejora de la calidad de los diagnósticos médicos con ultrasonido, sin el requerimiento de grandes volúmenes de imágenes de entrenamiento de los métodos de análisis basados en redes neuronales profundas. La combinación de las dos aproximaciones tiene potencial para hacer más efectivo el diagnóstico clínico con imágenes de ultrasonido en 2D y 3D, fijas y en movimiento.

Agradecimientos

Agradecemos el apoyo de la UNAM a través de los proyectos PAPIIT TA101121 y PAPIIT IV100420. Así como el apoyo de la SECTEI de la Cd. de México a través del proyecto SECTEI/202/2019.

BIBLIOGRAFÍA

- [1] Y. Yu & S. T. Acton, "Speckle reducing anisotropic diffusion," en *IEEE Transactions on image processing*, vol. 11, no. 11, pp.1260-1270, 2002.
- [2] K. Y. Leung, "Applications of Advanced Ultrasound Technology in Obstetrics," en *Diagnostics*, vol. 11, no. 7, p. 1217, 2021.
- [3] A. L. Klibanov & J. A. Hossack, "Ultrasound in radiology: from anatomic, functional, molecular imaging to drug delivery and image-guided therapy," en *Investigative radiology*, vol. 50, no. 9, p. 657, 2015.
- [4] T. Cootes, C. Taylor, D. Cooper, *et al.*, "Active shape models-Their training and application", en *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, 1995.
- [5] S. H. Contreras Ortiz, T. Chiu, and M. D. Fox, "Ultrasound image enhancement: A review," en *Biomedical Signal Processing and Control*, vol. 7, pp. 419–428, Sep 2012.
- [6] F. Liu, Z. Zhang, X. Lin, *et al.*, "Development of the human fetal cerebellum in the second trimester: a postmortem magnetic resonance imaging evaluation," en *Journal of anatomy*, vol. 219, no. 5, pp. 582-588, 2011.
- [7] M. R. Hatab, S. W. Kamourieh, & D. M. Twickler, "MR volume of the fetal cerebellum in relation to growth," en *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, no. 4, pp. 840-845, 2008.
- [8] G. Velásquez-Rodríguez, F. Arámbula-Cosío, & Boris-Escalante, "Automatic segmentation of the fetal cerebellum using spherical harmonics and gray level profiles," en *11th International Symposium on Medical Information Processing and Analysis*, vol. 9681, International Society for Optics and Photonics, 2015.
- [9] I. Timor-Trisch, A. Monteagudo, and H. Cohen, *Neuroecografía Prenatal y Neonatal*, vol. 1. MARBAN, 2004.
- [10] G. Kossoff, K. Griffiths, and C. Dixon, "Is the quality of transvaginal images superior to transabdominal ones under matched conditions?," *Ultrasound Obstet Gynecol*, vol. 1, pp. 29–35, Jan 1991.
- [11] J. Perez-Gonzalez, F. Arámbula-Cosío, J.C. Huegel, *et al.*, "Probabilistic Learning Coherent Point Drift for 3D Ultrasound Fetal Head Registration," en *Computational and Mathematical Methods in Medicine*, vol. 2020, pp. 1-14. <https://doi.org/10.1155/2020/4271519>.

- [12] J. Perez-Gonzalez, N. Hevia Montiel, V. Medina-Bañuelos, “Deep Learning Spatial Compounding from Multiple Fetal Head Ultrasound Acquisitions” en *International Workshop on Preterm, Perinatal and Paediatric Image Analysis, Held in Conjunction with MICCAI 2020*, Lima Perú. Lecture Notes in Computer Science, vol. 12437. Springer, Cham. https://doi.org/10.1007/978-3-030-60334-2_30.
- [13] J. Olveres, *et al.*, “What is new in computer vision and artificial intelligence in medical image analysis applications,” en *Quantitative Imaging in Medicine and Surgery*, vol. 11, no. 8, pp. 3830853–3833853, Aug. 2021, doi: 10.21037/QIMS-20-1151.
- [14] P. Jiang, J. Peng, G. Zhang, *et al.*, “Learning-based automatic breast tumor detection and segmentation in ultrasound images,” *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, May 2012, pp. 1587–1590, doi: 10.1109/ISBI.2012.6235878.
- [15] F. Torres, B. Escalante-Ramírez, J. Olveres, *et al.*, “Lesion Detection in Breast Ultrasound Images Using a Machine Learning Approach and Genetic Optimization,” In: Morales A., Fierrez J., Sánchez J., Ribeiro B. (eds) *Pattern Recognition and Image Analysis. IbPRIA 2019*. Lecture Notes in Computer Science, vol 11867. Springer, Cham. https://doi.org/10.1007/978-3-030-31332-6_26
- [16] A. Madabhushi and D. N. Metaxas, “Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions,” en *IEEE Transactions on Medical Imaging*, vol. 22, no. 2, pp. 155–169, 2003.
- [17] B. Liu, H. D. Cheng, J. Huang, *et al.*, “Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images,” en *Pattern Recognition*, vol. 43, no. 1, pp. 280–298, Jan. 2010, doi: 10.1016/J.PATCOG.2009.06.002.
- [18] F. Torres, Z. Fanti, and F. Arámbula-Cosío, “3D freehand ultrasound for medical assistance in diagnosis and treatment of breast cancer: Preliminary results,” in *Proceedings of SPIE - The International Society for Optical Engineering*, 2013, vol. 8922, doi: 10.1117/12.2041806.
- [19] C. Muñoz-Meza and W. Gomez, “A feature selection methodology for breast ultrasound classification,” en *2013 10th International Conference on Electrical Engineering, Computing Science and Automatic Control, CCE 2013*, pp. 245–249, 2013, doi: 10.1109/ICEEE.2013.6676056.
- [20] J. Shan, H. D. Cheng, and Y. Wang, “A novel automatic seed point selection algorithm for breast ultrasound images,” in *2008 19th International Conference on Pattern Recognition*, Dec. 2008, pp. 1–4, doi: 10.1109/ICPR.2008.4761336.

Cómo se cita:

F. Arámbula-Cosío, F. Torres-Robles, G.A.R. Velásquez-Rodríguez, *et al.*, “Aprendizaje computacional para análisis de imágenes de ultrasonido médico,” *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 4, octubre, 2021. [En línea]. Disponible en: <http://ties.unam.mx/> [Consultado en mes día, año].

Fecha de recepción: agosto 10, 2021

Fecha de publicación: octubre de 2021

APRENDIZAJE COMPUTACIONAL APLICADO A LA DETECCIÓN DE HUESOS, EN CIRUGÍA ORTOPÉDICA ASISTIDA POR COMPUTADORA

Zian Fanti, Alfonso Gastelum Strozzi, Eric Hazan Lasri y F. Arámbula Cosío
<http://www.ties.unam.mx/>

Fecha de recepción: agosto 10, 2021 • Fecha de publicación: octubre de 2021

Octubre 2021 | número de revista 4 • ISSN en trámite



APRENDIZAJE COMPUTACIONAL APLICADO A LA DETECCIÓN DE HUESOS, EN CIRUGÍA ORTOPÉDICA ASISTIDA POR COMPUTADORA

Resumen

En este trabajo presentamos la aplicación de un método de aprendizaje estadístico tomando como base el clasificador de Bayes, para detectar automáticamente la superficie de los huesos en imágenes de ultrasonido. La detección transoperatoria de los huesos del paciente permite el uso de modelos gráficos preoperatorios, de alta resolución, para guiar al cirujano durante la realización de un procedimiento ortopédico. Como caso de estudio reportamos el análisis de imágenes de ultrasonido y la construcción del modelo preoperatorio de la tibia. La distancia de error media, en la detección de automática de la superficie de la tibia, fue de 0.21 mm con una desviación estándar de 0.17 mm.

Palabras clave:

segmentación, registro, huesos, ultrasonido.

MACHINE LEARNING APPLIED TO BONE DETECTION IN COMPUTER ASSISTED ORTHOPEDIC SURGERY

Abstract

In this work is reported the application of a statistical learning method – Bayes classifier- to the automatic detection of the bone surface in ultrasound images. The intraoperative detection of the bone surface allows for the use of preoperative high resolution graphics models to guide the surgeon during an orthopedic procedure. As a study case we report the computer analysis of a 3D volume of ultrasound images of the tibial bone and the construction of the corresponding graphics model. The mean error distance of the detection of the tibial bone surface was 0.21mm with a standard deviation of 0.17 mm.

Keywords:

segmentation, registration, bones, ultrasound.

APRENDIZAJE COMPUTACIONAL APLICADO A LA DETECCIÓN DE HUESOS, EN CIRUGÍA ORTOPÉDICA ASISTIDA POR COMPUTADORA

I. Introducción

Las imágenes de ultrasonido se utilizan como la principal forma de imagenología en diversas especialidades médicas, en su mayoría para diagnosticar padecimientos en tejidos blandos como riñones o hígado. También es usado para dar seguimiento al desarrollo fetal durante el embarazo. Recientemente, se han utilizado para visualizar el tejido intraarticular, el cartílago, los ligamentos; y para diagnosticar la fractura de huesos que, en algunos casos puede sustituir a los tradicionales rayos X. [1] Otra alternativa viable, es la navegación y guía transoperatoria en cirugías ortopédicas asistidas por computadora (CAOS por sus siglas en inglés), donde las imágenes de fluoroscopia o de tomografía computarizada son las más usadas, esto debido a los beneficios que el ultrasonido brinda como: la portabilidad, el bajo costo, así como la posibilidad de adquirir imágenes libres de radiaciones.

En muchos procedimientos quirúrgicos, no basta con hacer una visualización de las imágenes de ultrasonido, es necesario contar con modelos anatómicos virtuales, para realizar una segmentación de los órganos de interés. En particular para este estudio serán los huesos, específicamente la meseta tibial.

La segmentación del hueso, y en general la segmentación de estructuras a partir de imágenes de ultrasonido, es una tarea difícil dadas las características físicas de este, principalmente por la baja relación entre señal y ruido.

En particular, las estructuras hiperecoicas que poseen los huesos (i. e. con alta densidad, que produce ecos ultrasónicos de alta amplitud), producen una gran sombra acústica, reverberaciones, y una respuesta distinta dependiendo del ángulo de incidencia de las ondas acústicas.

En este artículo reportamos la segmentación de la superficie ósea de la meseta tibial. Se implementó el aprendizaje de máquina basado en un clasificador supervisado de Bayes, con el objetivo de demostrar la viabilidad del uso de esta metodología en la segmentación de volúmenes de ultrasonido, los cuales se obtuvieron mediante la técnica conocida como “ultrasonido 3D a manos libres” descrita en *Image tracking and volume reconstruction of medical ultrasound*. [2] El objetivo final es poder realizar el registro transoperatorio en una cirugía ortopédica asistida por computadora, con el modelo virtual y el plan preoperatorio del paciente, como se reportó en *Improved surface-based registration of CT and intraoperative 3D ultrasound of bones*. [3]

II. Detección de la superficie ósea en ultrasonido 3D

Las estructuras óseas aparecen en las imágenes de ultrasonido como regiones con bordes brillantes seguidas de una región oscura, resultado de la sombra acústica generada por el hueso como se observa en la figura 1. La alta intensidad del reflejo acústico provocado por el hueso se observa en las imágenes como una línea de espesor

entre 2 y 4mm. Característica que debe ser tomada en cuenta para que las dimensiones de la superficie extraída correspondan con el tamaño original del hueso del paciente. Jain *et al.*, [4] demuestran que la cortical del hueso se encuentra en el máximo de intensidad del gradiente generado por la respuesta del hueso al ultrasonido. El resultado activa un detector que realiza la línea media para que tenga una representación más precisa de la cortical ósea en imágenes de ultrasonido que los métodos basados en detección de bordes.

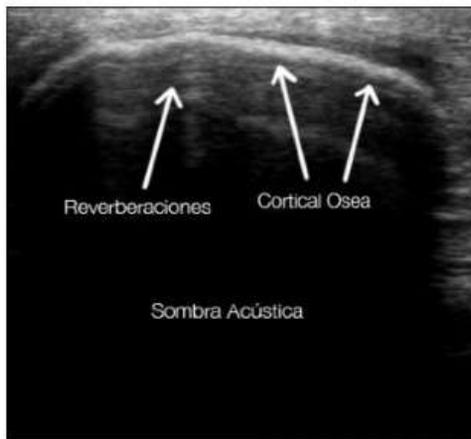


Figura 1. Muestra una imagen de ultrasonido de la meseta tibial. En esta se pueden apreciar artefactos del ultrasonido como la sombra acústica, las reverberaciones y la cortical ósea

Nuestro método de segmentación toma en cuenta alteraciones como el ruido *speckle* y las reverberaciones, para minimizarlos se desarrollaron dos procesos: el realce de estructuras óseas y la posterior segmentación sobre la imagen realzada, como se describe a continuación.

II.1. Realce de superficies en ultrasonido 3D

La ubicación de una cresta local en una función se da cuando la segunda derivada es igual a cero. Sea I un volumen de ultrasonido, este se puede expresar como una función en tres dimensiones multivariable, por lo que las segundas derivadas direccionales en x , y y z del volumen describen la variación del gradiente de intensidad de los voxeles localmente. La segunda derivada direccional en un punto $\hat{x}=(x,y,z)$ del volumen se representa como matriz hessiana (Ec. 1, apéndice 1), donde cada elemento representa la derivada parcial de segundo orden de la

imagen en cada dirección (x,y,z) . Una propiedad de esta matriz es que es simétrica con eigenvalores reales y eigenvectores ortogonales, por lo tanto, invariante a rotaciones.

La matriz hessiana se construye a partir del cálculo de la segunda derivada parcial direccional, que puede ser aproximada mediante la convolución del volumen I con la segunda derivada de la función Gaussiana normalizada G (Ec. 2 y 3, apéndice 1), donde la desviación estándar σ de G es usado como un factor de escala, el cual se asignará dependiendo del tamaño en la imagen de las estructuras que se quieran realzar.

La segunda derivada de la imagen aplicada sobre un voxel, proporciona información del tipo de estructura a la que pertenece, pueden ser formas tubulares, esféricas o de superficie. Esta información se obtiene del eigenanálisis de la matriz hessiana generada para cada voxel en el volumen (Tabla 1, apéndice 1). Se eligen solo los voxeles que cumplan con la condición de pertenecer a una superficie, un valor de "superficialidad" les es asignado (Ec. 4, apéndice 1) y como resultado se obtiene una imagen 3D donde cada voxel representa una medida de "superficialidad" para una escala σ dada, la cual llamaremos imagen realzada.

II.2. Clasificación de las superficies de los huesos

Posteriormente a partir de la imagen realzada, la segmentación de la cortical del hueso se hace usando un clasificador supervisado de Bayes, el cual asigna cada voxel en una de las posibles clases del conjunto $C=c_1, c_2, \dots, c_m$ definidas previamente. La asignación de un voxel a una clase se da cuando la probabilidad de pertenecer a la clase C_k dado un vector que caracteriza a este, (\hat{X}) es máxima sobre las demás clases. Esta probabilidad se obtiene a partir del teorema de Bayes definido en la ecuación 6 en el apéndice 1.

Si asumimos que la distribución N -dimensional para cada una de las clases es Gaussiana, se puede obtener el discriminante de Bayes (Ec. 7, apéndice 1) para cada una de las clases k , el cual representará la probabilidad del vector \hat{X} de pertenencia a la clase k .

Para las imágenes de ultrasonido de hueso se definieron tres clases, que corresponden a diferentes regiones y claramente distintas entre sí: el tejido, que es todo lo visible por encima del hueso; la cortical ósea que se observa como la línea más brillante en la imagen; y la sombra acústica, que es la sección oscura por debajo del hueso. El vector de características para cada voxel se definió como: la intensidad del voxel en el volumen realzado

$(E(x,y,z))$, y los tres primeros momentos estadísticos del volumen realzado ($M_{e1}(x,y,z)$, $M_{e2}(x,y,z)$, $M_{e3}(x,y,z)$) definidos en las ecuaciones 8, 9 y 10 respectivamente, incluidas en el apéndice 1.

Una vez construido el vector de características para cada voxel en el volumen se procede a entrenar el clasificador supervisado, para esto se necesita un volumen de entrenamiento que tenga tres conjuntos de voxeles etiquetados por un radiólogo experto con la clase a la que pertenecen. Con este volumen se encuentran la media (μ_k) y la matriz de covarianza (Σ_k) para cada clase k .

El entrenamiento del sistema solo se realiza una vez para imágenes pertenecientes a un mismo protocolo de adquisición y se puede realizar fuera de línea. En la etapa de clasificación, se construye una imagen de probabilidad $I_p(x,y,z)$, donde el valor de cada voxel corresponde a la probabilidad de pertenecer a la clase hueso. Si el valor de probabilidad es mayor para otra clase se le asignará a dicho voxel un valor de cero. Finalmente, a partir de la imagen de probabilidad se selecciona un porcentaje mínimo de voxeles con el valor de probabilidad más alto de pertenecer a la cortical ósea, a este conjunto se designará como el conjunto de semillas, las cuales son usadas para inicializar un algoritmo de crecimiento de regiones, [5] a lo largo de todos los voxeles vecinos cuyo valor de probabilidad es $I_p > \epsilon$. Donde ϵ es un valor que se asigna de forma experimental con el cual podemos modular que tan “grosso” se quiere la sección de la cortical del hueso. En esta parte, el algoritmo de crecimiento de regiones se usa para evitar que artefactos como las reverberaciones sean clasificadas como hueso, ya que estas presentan una característica similar.

II.3. Adquisición de imágenes de ultrasonido 3D de la tibia

Para validar el método de segmentación se utilizó un maniquí realista, que corresponde a la tibia como se muestra en la figura 6. Se adquirieron 12 volúmenes de ultrasonido de distintas partes que involucran a la meseta tibial. Las imágenes de ultrasonido del maniquí se adquirieron con la técnica de ultrasonido navegado 3D de mano libre descrita en Torres *et al.*, [2] En la figura 2 se observan tres planos ortogonales de ultrasonido del maniquí de la tibia.

Se usó un modelo virtual de imágenes de tomografía computarizada del maniquí, para evaluar los resultados de la segmentación de la superficie de la tibia en el volumen de ultrasonido, como se describe a continuación.

II.4. Modelo virtual de tomografía computarizada (TC) de la tibia

Uno de los métodos de tomografía más utilizados son los métodos de tomografía computarizada por rayos X (TC) como lo señalan Fishman & Jeffrey. [6] La TC funciona obteniendo imágenes de secciones del objeto en forma secuencial. Existen diversas configuraciones para la adquisición de estas secciones, por un lado, se cuenta con la configuración helicoidal utilizada comúnmente en la medicina, como se muestra en la figura 3, y la utilizada en la industria, la cual utiliza una mesa giratoria como el sistema de adquisición en los sistemas de microtomografía, figura 4. [7]

Los TC proveen un grupo de imágenes I , donde cada una representa una sección 2D del volumen, el grupo de todas define el volumen digitalizado $V \in \mathbb{Z}^3$. Al igual que con las imágenes donde el concepto de píxel [8] se utiliza, para definir la menor unidad homogénea que contiene in-

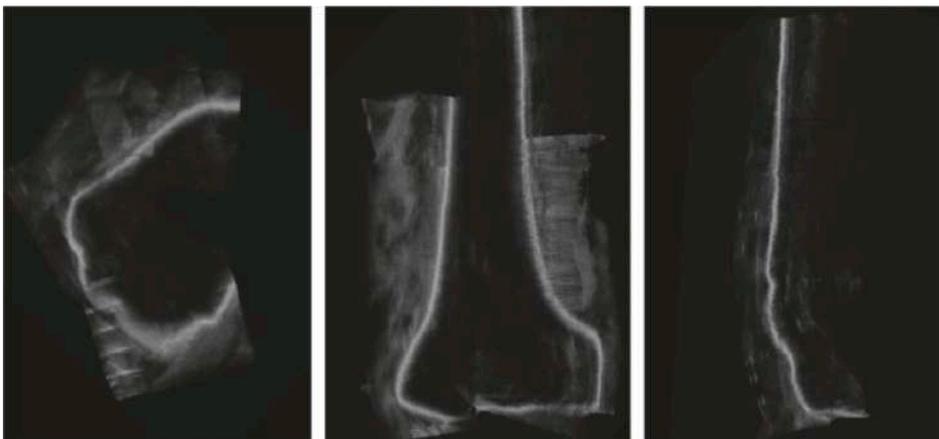


Figura 2. Planos ortogonales de ultrasonido 3D de un maniquí de la tibia

formación de la imagen. Los volúmenes V están formados por voxels $Y \in V$ y son los elementos procesables más pequeños. La resolución espacial de nuestro volumen depende de la resolución del sensor y la distancia recorrida entre cada toma, es común tener diferentes resoluciones en cada dirección espacial.

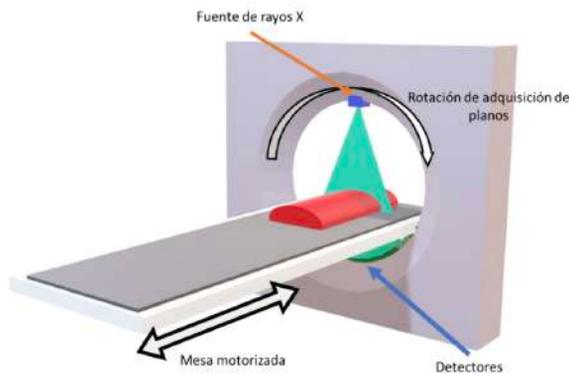


Figura 3. Configuración helicoidal para la obtención de imágenes tomográficas. La mesa motorizada y el conjunto rotacional de rayos X y sensores se mueven de forma sincronizada para adquirir el volumen 3D

representar varias radiodensidades de forma discreta. Por ejemplo, un sensor de 8 Bits puede guardar 256 diferentes radiodensidades como valores de gris. Y la complejidad de los materiales a adquirir en una toma define que porciones del rango dinámico son usadas para representar cada material. En la figura 5 se muestra una imagen obtenida por técnicas de CT con su rango dinámico. [10]

Para construir el modelo virtual de la tibia se usó un modelo de CT basado en un maniquí de tibia. Este fue construido con hueso tibial sintético (ERP #1117-42, Sawbones Inc., Vashon, WA, I.E.E.U.U.) sumergido en un hidrogel hecho del alcohol de polivinilo (PVA), diluido en el 95% de agua. En la figura 6 se muestra una imagen del hueso sintético y del maniquí final cubierto de hidrogel.

El maniquí fue digitalizado con un MicroCT, modelo Nikon Metrology XTH 225, con resolución de imagen de 2048 x 2048. Se configuró el sistema, para obtener imágenes utilizando 220 kV y 61 μ A, con un número de proyecciones de 3142. Esta configuración permitió

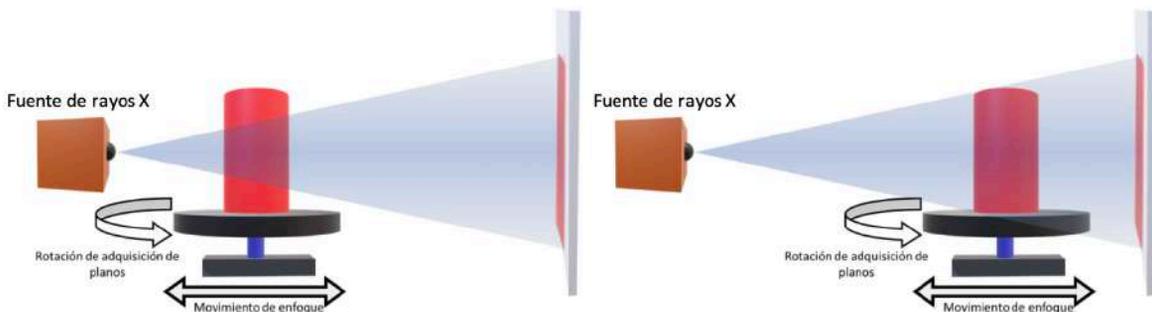


Figura 4. Configuración clásica de un micro tomógrafo. En estos sistemas la mesa rotatoria gira, para adquirir un plano (imagen) por ángulo. Un segundo sistema de posicionamiento (movimiento de enfoque) permite seleccionar el volumen de interés a ser adquirido

Los valores guardados en cada voxel y el proceso en el que la imagen tomográfica se forma pueden entenderse como un símil a la transparencia y translucencia de objetos con respecto a la luz visible, donde los materiales presentan diferentes valores de radiodensidad que indican que tan opaco es el material a los rayos X como lo señalan Novelline & Squiere. [9]

La profundidad de bits [10] del sensor y los materiales del objeto a adquirir definen el rango dinámico del volumen, cada sensor tiene distintas capacidades para

obtener un volumen final compuesto por $1042 \times 1250 \times 3201$ voxels donde cada voxel tiene una resolución de 0.115 mm^3 .

La superficie ósea se obtuvo etiquetando el voxel que contiene información del hueso sintético con el valor de 1 y el resto de los voxels se les asignó un valor de 0. El proceso de etiquetado se realizó utilizando un algoritmo de agrupamiento k-means. [11] El volumen del hueso obtenido del CT se muestra en la figura 7.

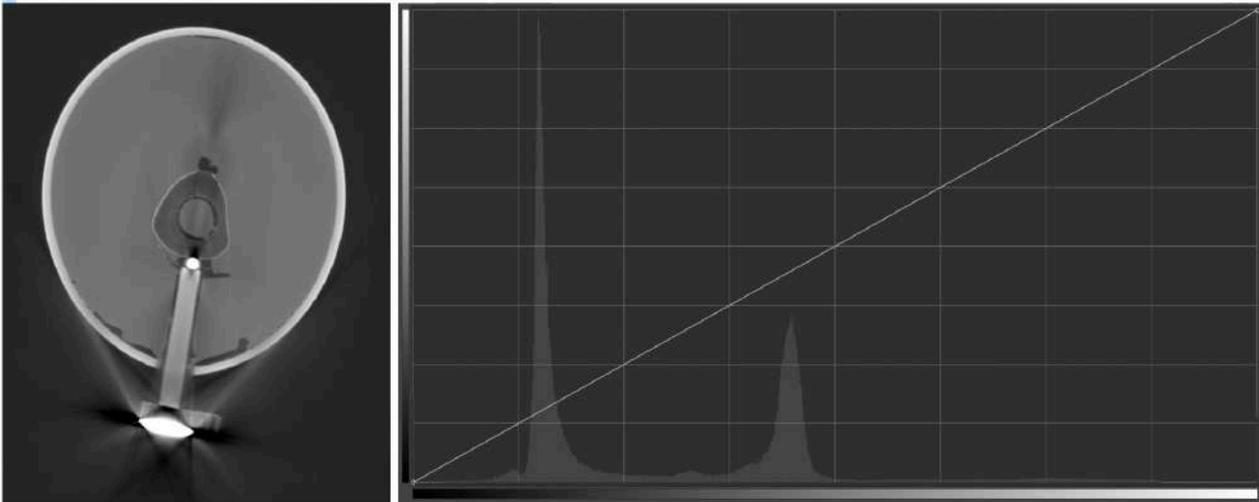


Figura 5. Imagen tomográfica mostrando una composición bimodal de intervalo dinámico. La primera moda contiene en su mayoría voxeles digitalizando aire. La segunda moda se debe al modelo del hueso, la carne, el contenedor y un objeto metálico



Figura 6. (arriba) hueso tibial sintético ERP #1117-42. (abajo) maniquí de hueso cubierto de PVA.

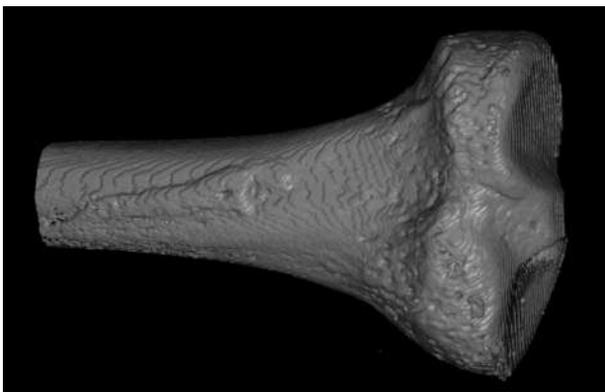


Figura 7. Modelo 3D del hueso digitalizado

III. Resultados

Para cada volumen segmentado a partir de las imágenes de ultrasonido se montó el modelo virtual de CT y se obtuvo la distancia mínima al volumen de referencia para cada uno de los voxeles segmentados. En promedio, se registró una distancia máxima de 2.031 mm, sin embargo, el promedio de la distancia media es 0.21 mm con una desviación estándar de 0.17 mm. Esto muestra que muy pocos de los voxeles segmentados están fuera de la cortical del hueso. En la figura 8 (sección superior) se muestra la imagen original de US, la segmentación de referencia en amarillo y la segmentación del volumen de US sobrepuesta en azul.

Una vez segmentado el volumen, y a partir de este, se construye una malla la cual representa la anatomía del paciente en la mesa de operaciones. El modelo preoperatorio se pone en el mismo sistema de referencias que la segmentación para poder iniciar el proceso de registro, con el algoritmo *Iterative Closest Points*. [12] La figura 8 (sección inferior) muestra el resultado del registro de las dos mallas, la segmentada en azul y la del plan preoperatorio en amarillo.

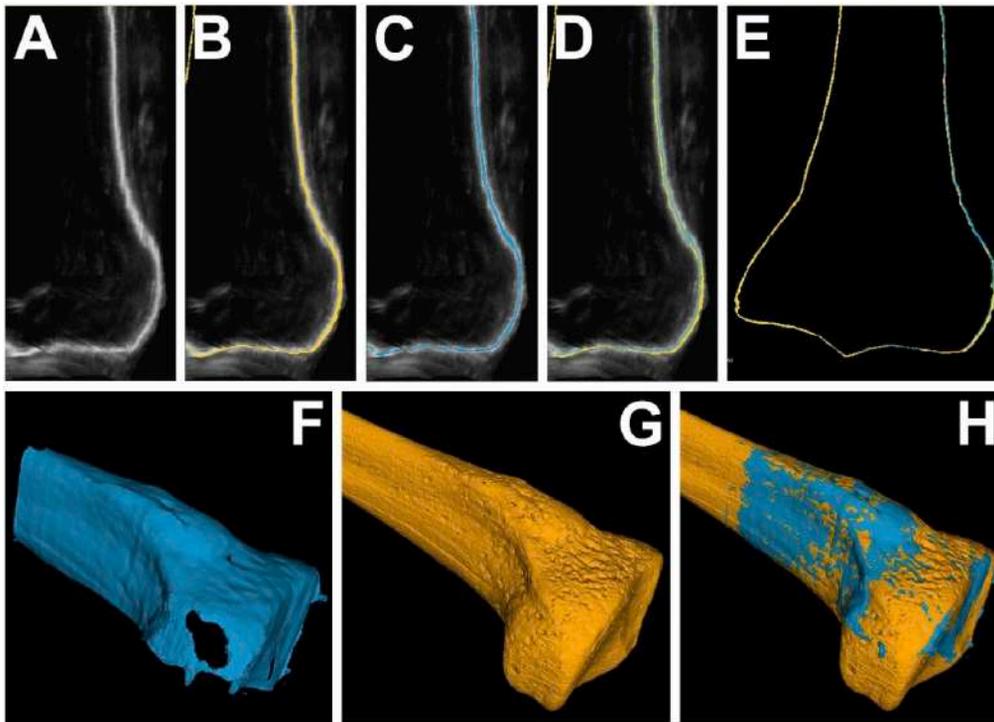


Figura 8. Muestra un ejemplo del resultado de la segmentación. En A) se muestra un corte del volumen de ultrasonido de la meseta tibial, B) el corte de A) con el modelo de validación sobrepuesto, C) el corte de A) con el resultado de la segmentación sobre puesto, D) el corte de A) con el modelo de validación y el resultado de la segmentación sobre puestos, E) El modelo de validación de TC y el resultado de la segmentación sobrepuestos, F) la reconstrucción 3D del resultado de la segmentación, F) el modelo de validación y H) el modelo virtual de validación y el resultado de la segmentación sobrepuestos.

IV. Conclusiones

En este trabajo se presenta un método, para la detección automática de la superficie de los huesos en imágenes de ultrasonido 3D, basado en la detección de estructuras 3D: superficies, líneas, esferas y en la clasificación automática de la superficie de los huesos utilizando un clasificador de Bayes. La validación del método de segmentación en el maniquí de tibia produjo resultados con errores de exactitud suficientemente pequeños como para implementar este método en la detección de huesos con fines quirúrgicos, aunque una validación extendida a otros huesos será necesaria.

Anexo 1

Configuración de la matriz hessiana:

$$H(\hat{x}) = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix} \quad (1)$$

cada elemento I_{ij} representa la derivada parcial de segundo orden de la imagen en la dirección ij en la posición \hat{x} , esto es, $I_{xx} = \frac{\partial^2 I}{\partial x^2}$, $I_{xy} = \frac{\partial^2 I}{\partial x \partial y}$,... y así sucesivamente. Cuando $i=j$ entonces $I_{ij}=I_{ji}$, esta propiedad hace que la matriz hessiana sea simétrica con eigenvalores reales y eigenvectores ortogonales, por lo tanto, invariante a rotaciones.

Las segundas derivadas se aproximan con la convolución del volumen I con la segunda derivada de la función Gaussiana normalizada G , de esta manera para la dirección ij , el elemento I_{ij} de la matriz hessiana se define como:

$$I_{ij}(i, j; \sigma) = I(i, j) \otimes \frac{\partial^2}{\partial i \partial j} G(i, j; \sigma) \quad (2)$$

donde:

$$G(i, j; \sigma) = \frac{1}{2\pi^2} e^{-\left(\frac{i^2+j^2}{2\sigma^2}\right)} \quad (3)$$

con σ el valor de la desviación estándar. El valor de σ es también utilizado como un factor de escala, el cual se asignará dependiendo del tamaño en la imagen de las estructuras que se quieran realzar.

El eigenanálisis para cada voxel en el volumen resulta en la obtención de tres eigenvalores λ_1, λ_2 y λ_3 . Sin pérdida de generalidad, sean $\lambda_1 \geq \lambda_2 \geq \lambda_3$, se puede encontrar un conjunto de condiciones discriminantes de pertenencia a un tipo de estructura, la cual puede ser tubular, esférica o de superficie. [6] Estas condiciones se muestran a continuación en la tabla 1:

| Estructura | Eigenanálisis |
|------------|---|
| Superficie | $\lambda_3 \ll \lambda_2 \approx \lambda_1 \approx 0$ |
| Línea | $\lambda_3 \approx \lambda_2 \ll \lambda_1 \approx 0$ |
| Esfera | $\lambda_3 \approx \lambda_2 \approx \lambda_1 \ll 0$ |

Tabla 1. Muestra las distintas configuraciones que deben de cumplir los eigenvalores de la matriz hessiana, para pertenecer a las estructuras tubulares, esféricas o de superficie

La medida de superficialidad para cada voxel se puede asignar a partir de la siguiente ecuación:

$$S_{superficie}(I) = \begin{cases} |\lambda_3| \cdot \omega(\lambda_2, \lambda_3) \cdot \omega(1, \lambda_3) & \text{si } \lambda_3 < 0 \\ 0 & \text{en otro caso} \end{cases} \quad (4)$$

donde:

$$\omega(\lambda_s, \lambda_t) = \begin{cases} \left(1 + \frac{\lambda_s}{|\lambda_t|}\right)^y & \text{si } \lambda_t \leq \lambda_s \leq 0 \\ \left(1 - \alpha \frac{\lambda_s}{|\lambda_t|}\right)^y & \text{si } \frac{|\lambda_t|}{\alpha} \leq \lambda_s \leq 0 \\ 0 & \text{en otro caso} \end{cases} \quad (5)$$

los parámetros α y γ fueron escogidos experimentalmente, obteniendo $\alpha=0.5$ y $\gamma=1$.

Con la imagen de probabilidades y basados en el teorema de Bayes se puede obtener la probabilidad de pertenecer a una clase k dado un vector de características:

$$C_k = \operatorname{argmax}_k \left(P(\hat{X}|C_k) \cdot P(C_k) \right) \quad (6)$$

donde:

$P(C_k)$ es la probabilidad a priori de ocurrencia para la clase k .

$P(\hat{X} | C_k)$ es la probabilidad condicional de \hat{X} dado C_k . \hat{X} el vector $u^T = u_1, u_2, \dots, u_n$ un que caracteriza un voxel del volumen.

Si asumimos que la distribución N-dimensional para cada una de las clases es Gaussiana, el discriminante de Bayes para cada una de las clases k se define como:

$$y_k(\hat{X}) = -\frac{1}{2}(\hat{X} - \mu_k)^T \Sigma_k^{-1}(\hat{X} - \mu_k) - \frac{1}{2} \ln |\Sigma_k| + \ln P(C_k) \quad (7)$$

donde:

A es la dimensión del vector \hat{X}

μ_k es un vector N-dimensional que representa la media de cada una de las características del conjunto de entrenamiento.

Σ_k es una matriz de $N \times N$ la matriz de covarianza para C_k . $|\Sigma_k|$ el determinante de la matriz de covarianza.

Los tres primeros momentos estadísticos se definen como:

$$M_{e1}(\hat{x}) = \frac{1}{n} \sum_{i=1}^n I(\hat{x}) \quad (8)$$

$$M_{e2}(\hat{x}) = \frac{1}{n} \sum_{i=1}^n (I(\hat{x}) - \mu)^2 \quad (9)$$

$$M_{e3}(\hat{x}) = \frac{1}{n} \sum_{i=1}^n (I(\hat{x}) - \mu)^3 \quad (10)$$

El cálculo de los momentos estadísticos se realizó dentro de una ventana de tamaño fijo escogido como tres veces el tamaño ocupado por la respuesta del hueso en el ultrasonido, esto con el fin de poder obtener la relación entre los voxeles pertenecientes a la cortical del hueso y los voxeles vecinos a esta.

BIBLIOGRAFÍA

- [1] A. L. Waterbrook, S. Adhikari, U. Stolz, *et al.*, "The accuracy of point-of-care ultrasound to diagnose long bone fractures in the ED." en *The American journal of emergency medicine*, vol. 31, no. 9, pp. 1352-1356. 2013.
- [2] [2] F. Torres, Z. Fanti, E. Lira, *et al.*, "Image tracking and volume reconstruction of medical ultrasound," en *Revista Mexicana de Ingeniería Biomedica*, vol. 33, no. 2, pp. 101-115, 2012.
- [3] Z. Fanti, E. Torres, A. Hazan-Lasri, *et al.*, "Improved surface-based registration of CT and intraoperative 3D ultrasound of bones," en *Journal of healthcare engineering*, 2018.
- [4] A. K. Jain, & R. H. Taylor, "Understanding bone responses in B-mode ultrasound images and automatic bone surface extraction using a Bayesian probabilistic framework," in *Medical imaging 2004: ultrasonic imaging and signal processing*, vol. 5373, pp. 131-142. International Society for Optics and Photonics. 2004, April.
- [5] R., Adams, & L. Bischof, "Seeded region growing," en *IEEE Transactions on pattern analysis and machine intelligence*, vol.16, no. 6, pp. 641-647, 1994.
- [6] E. K. Fishman, & R. B. Jeffrey, *Spiral CT: Principles, Techniques, and Clinical Applications*. Raven Press, 1995.
- [7] E. L. Ritman, "Current Status of Developments and Applications of Micro-CT," en *Annual Review of Biomedical Engineering*, vol. 13, no. 1, pp.531-552, 2011.
- [8] J. D. Foley, & A. Van Dam, *Fundamentals of interactive computer graphics*. Reading, Mass, Addison-Wesley Pub. Co, 1983.
- [9] [9] R. A. Novelline, & L. F. Squire, *Squire's Fundamentals of Radiology*. La Editorial, UPR, 2004.
- [10] "An In-Depth Look at Bit Depth," *Blog Post Teledyne Lumenera*. (n.d.). [En línea]. Disponible en <https://www.lumenera.com/blog/bit-depth> [Consultado en august 9, 2021].
- [11] G. A. F. Seber, *Multivariate Observations*. John Wiley & Sons, 2004.
- [12] P. J. Besl, & N. D. McKay, "Method for registration of 3-D shapes. In Sensor fusion IV: control paradigms and data structures," en *International Society for Optics and Photonics*. vol. 1611, pp. 586-606, 1992, April.

Cómo se cita:

Z. Fanti, A. Gastelum-Strozzi, E Hazan-Lasri, *et al.*, “Aprendizaje computacional aplicado a la detección de huesos, en cirugía ortopédica asistida por computadora,” *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 4, octubre, 2021. [En línea]. Disponible en: <http://ties.unam.mx/> [Consultado en mes día, año].

Fecha de recepción: agosto 10, 2021

Fecha de publicación: octubre de 2021