

Desarrollo de un conjunto de datos de imágenes de hojas de batata como herramienta para reconocimiento de diferentes genotipos

Antonella Vilanova Perez¹, Sofía Solange Flamarique¹,
Liliana del Valle Di Feo¹ y Javier Berger²

¹ Instituto de Patología Vegetal (IPAVE), Centro de Investigaciones Agropecuarias (CIAP),
Instituto Nacional de Tecnología Agropecuaria (INTA)
Camino 60 cuadras km 5.5, Córdoba, Argentina
{vilanova.antonella;difeo.liliana}@inta.gob.ar,
soflamarique@gmail.com

² Universidad Nacional de Misiones, Facultad de Ciencias Exactas, Químicas y Naturales
Félix de Azara N° 1552, Posadas (3300), Misiones, Argentina
javiberger@gmail.com

Resumen. Un conjunto de datos de imágenes proporciona una fuente muy importante y necesaria para la realización de diversas actividades de investigación y análisis en varios campos computacionales relacionados con la inteligencia artificial como son la visión por computadora, el aprendizaje automático y el aprendizaje profundo. En el presente trabajo, se proporciona un conjunto de datos de imágenes de hojas del cultivo de batata de cuatro variedades morfológicamente contrastantes y ampliamente cultivadas en el país. Esta contribución incluye 836 imágenes digitales y se pone a disposición pública, especialmente para la comunidad de investigadores, bajo licencia Creative Commons con Atribución.

Palabras clave: conjunto de datos, visión artificial, inteligencia artificial, *Ipomoea batatas*

1 Introducción

La batata, *Ipomoea batatas* (L.) Lam., hierba perenne convolvulácea cultivada como anual en zonas templadas [1] y originaria del noroeste de Sudamérica [2] [3] se ubica entre las 10 especies vegetales más importantes destinadas a alimentación, con una producción cercana a 91.945.358 t/año en 8.062.737ha plantadas en el mundo [4].

Las áreas más extensas de producción se encuentran en China y rodeando los Grandes Lagos del este de África, pero el cultivo también es relevante en otros países [5] [6]. Sin embargo, en países en desarrollo los rendimientos son bajos, pese a ser la tercera especie de importancia entre las raíces y tubérculos, luego de papa y yuca [4].

Presenta amplia adaptabilidad y versatilidad de usos, merced a su gran diversidad genética y consiguiente variabilidad en sus características fenotípicas y morfológicas. Debido a su alto rendimiento, bajo costo de producción y alto valor nutricional (hasta 10% de proteínas, alto contenido de almidón, fibras, minerales, vitaminas y antioxidante) se considera un cultivo clave para la seguridad alimentaria y nutricional en diversas regiones del mundo [7] [8]. El uso de variedades con pulpa amarilla o anaranjada (alto contenido de β -carotenos) está logrando disminuir la incidencia de la ceguera infantil en África.

Además, posee un mercado potencial para la exportación en fresco (Canadá, Inglaterra, Países Bajos, Suecia y otros) de aproximadamente 200 mil t/año y se prevé un incremento en el procesamiento de batata para extraer almidón debido a sus múltiples posibilidades de uso industrial.

En Argentina y teniendo en cuenta las actuales tendencias en los hábitos alimentarios saludables, la batata posee demanda culinaria creciente, por lo que podría convertirse en un producto de mayor participación en la dieta. La producción actual es de unas 339.359 t/año en 22.772ha plantadas, siendo el consumo de 3 kg/hab/año [9] [4]. La región pampeana (Buenos Aires, Córdoba y Santa Fe) y el NEA poseen 43 y 40% de superficie plantada, respectivamente; el NOA, 15% y Cuyo, el 2% restante [10].

Debido a la importancia de este cultivo, el Instituto Nacional de Tecnología Agropecuaria (INTA), lleva adelante diversas líneas de investigación sobre este cultivo, contempladas en la cartera de proyectos 2019.

En el Instituto de Patología Vegetal (IPAVE) dependiente del Centro de Investigaciones Agropecuarias (CIAP) del INTA, además, desde hace una década se producen, multiplican y distribuyen propágulos de sanidad controlada a productores, técnicos, cooperativas y distintos actores de la cadena de esta hortícola. Esta institución posee una colección de genotipos de batata que satisface las preferencias de las distintas zonas de producción del país. El reconocimiento de los mismos por simple observación visual de su parte aérea es bastante dificultoso, debido a la gran variabilidad genética de la especie que conlleva diversidad morfológica foliar y, por otro lado, muchos cultivares poseen hojas con características muy similares, pese a diferencias notables en sus raíces.

Por tal motivo, la identificación de variedades de batata que se plantan en las diferentes regiones de cultivo, requiere que las personas encargadas de esta tarea cuenten con adecuado entrenamiento visual. El desarrollo reciente de tecnologías vinculadas con el campo de la inteligencia artificial produjo herramientas de soporte, como aplicaciones para dispositivos móviles, que tienen como objetivo ayudar a las personas a identificar las variedades de un cultivo específico mediante el uso de fotografías previamente etiquetadas visualmente por un experto. Ejemplos de esto son la identificación de cultivares de mandioca [11], la clasificación de variedades de trigo [12] y el reconocimiento de variedades de manzana [13].

La mayoría de los modelos de clasificación de imágenes se encuentran disponibles en repositorios de la web, pero para utilizarlos es necesario contar con un conjunto de

datos etiquetados y los mismos son complejos de obtener, más aún, es muy difícil certificar que las etiquetas asignadas son realmente las que representan al elemento, como por ejemplo una fotografía.

Por esta razón ofrecer un conjunto de imágenes etiquetadas es fundamental para cooperar con el desarrollo de tecnologías que beneficien a todos los actores del sector productivo, ya que estos elementos son la base de la elaboración de modelos de clasificación. Así, un conjunto de datos de imágenes de cultivares de batata etiquetadas por expertos de IPAVE, un centro de investigación referente en la región para este cultivo, brinda el soporte necesario al desarrollo de modelos de aprendizaje profundo útiles para clasificar cultivares a través de fotografías de sus hojas.

Este trabajo propone la ejecución de estudios preliminares que, a partir del conjunto de datos de imágenes de hojas, permitan reconocer con alto grado de precisión el genotipo de batata al que pertenecen.

2 Reconocimiento de genotipos de batata

El reconocimiento de genotipos de batata mediante el uso de fotografías requiere de la aplicación de métodos de visión artificial o visión por computadora. Este conjunto de tecnologías corresponde a las ciencias de la computación, específicamente al ámbito de la inteligencia artificial. Un sistema de visión artificial involucra la adquisición, procesamiento, clasificación y reconocimiento de imágenes digitales [14].

La primera fase consiste en la adquisición de las imágenes relacionadas con la problemática que se intenta resolver mediante el sistema de visión artificial. Este conjunto de imágenes se puede obtener de un repositorio de datos o se puede construir mediante equipos de digitalización como ser cámara digitales o escáneres. En el caso que sea necesario construir el conjunto de datos es necesario tener en cuenta aspectos como la iluminación a utilizar, el fondo de las imágenes y el ángulo de adquisición para que estos aspectos faciliten el posterior procesamiento y clasificación de las imágenes. Además, un aspecto fundamental, es la fuente de la que provienen los datos ya que es necesario certificar que las etiquetas otorgadas a cada imagen provengan de expertos para evitar construir clasificadores que generen errores al momento de llevar la solución al campo de trabajo.

En la etapa de procesamiento se aplican distintas técnicas que tienen como objetivo mejorar la calidad de la imagen para facilitar el posterior reconocimiento y clasificación en base a patrones detectables en los píxeles que conforman la imagen. Entre las técnicas más importantes se encuentra la de detección y realzado de bordes mediante la identificación de cambios bruscos en la intensidad de los píxeles. Como así también, la técnica de eliminación de ruido que tiene como objetivo la supresión de píxeles cuyo nivel de intensidad difiera con la intensidad de los píxeles vecinos y, también, la reducción de variaciones de intensidad de los píxeles mediante la técnica de suavizado.

Finalizada la fase de procesamiento se ejecuta la etapa de clasificación y reconocimiento de las imágenes. En la misma se pueden aplicar diferentes métodos como son los árboles de decisión, naive bayes, máquinas de soporte vectorial, redes neuronales

artificiales, entre otros. Cada clasificador posee ventajas y desventajas que lo hace adecuado para encarar diferentes problemáticas de reconocimiento de imágenes. El avance alcanzado con los distintos clasificadores acarrió que en el presente los métodos relacionados con el aprendizaje automático, específicamente con el campo del *deep learning* o aprendizaje profundo se destaquen en la clasificación de imágenes por el nivel de acierto que poseen los modelos de redes neuronales convolucionales que integran estos métodos [15].

Por lo expresado, el presente trabajo se enfoca en la fase inicial del desarrollo de un sistema para el reconocimiento de genotipos de batata. Esta etapa se considera fundamental ya que de la misma se desprenden las demás. Por esta razón, es necesario que los datos adquiridos tengan la calidad necesaria para garantizar el correcto funcionamiento de las técnicas de procesamiento y los métodos de clasificación de imágenes [16].

3 Contexto de recopilación de datos

Las imágenes recolectadas y presentadas en este trabajo provienen de plantas de batata con sanidad controlada del IPAVE-CIAP-INTA.

El IPAVE cuenta, al presente, con una colección de 18 genotipos de batata requeridos en las diversas provincias productoras, que fueron saneados de virus y de otros patógenos sistémicos a través del cultivo *in vitro* de meristemas con termoterapia previa, y que recuperaron su pureza varietal. Los mismos son mantenidos por micropropagación *in vitro* y macropropagación en invernaderos y jaulón antiáfidos.

Actualmente, la institución ofrece a los agricultores y a la agroindustria propágulos de sanidad controlada de los cultivares Arapey INIA, Beauregard, Covington, Gem, Jewel, Morada INTA, clon Mechada, Boni INTA, Okinawa 100, Colorado INTA, Sombrerito, Pionera, Bonita, Paraguaya y Famaillá 6, para su distribución en las regiones productoras argentinas, además de tres clones promisorios aún no descriptos, de interés local.

El set de datos presentado en este trabajo se compone de 836 imágenes de hojas de batata de los cultivares Arapey INIA (raíces con pulpa crema con inclusiones anaranjadas y piel morada); Beauregard (raíces con pulpa completamente anaranjada y piel rosada cobriza); Boni INTA (pulpa y piel amarillo intenso) y Morada INTA (pulpa amarillo cremosa con inclusiones anaranjadas y piel morada).

La cara adaxial o haz y la abaxial o envés de distintas hojas, fueron fotografiadas en cada genotipo: Arapey INIA 56 fotos del haz y 61 del envés; Beauregard 130 y 115; Boni INTA 133 y 133 y Morada INTA, 113 y 95, respectivamente (Fig.1).

Las plantas de cada genotipo utilizado son clones, es decir, todas provienen de una única vitroplanta obtenida a partir de un meristema regenerado bajo condiciones controladas, aclimatado luego en invernadero y multiplicado en módulos de producción a campo bajo protección con malla antiáfidos.

Las fotos fueron tomadas a campo con un teléfono celular Samsung Galaxy J5 Prime, modelo: SM-G570M (2015) en el caso de Arapey INIA, y con el mismo celular y el celular Samsung Galaxy S9, modelo SM-G9600 (2018) para Beauregard, Boni INTA y Morada INTA. Las imágenes se entregan en formato jpg y las dimensiones varían entre 2322 x 4128 píxeles para el dispositivo Samsung Galaxy J5 Prime y 3024 x 4032 píxeles para el Samsung Galaxy S9, ambos con resolución horizontal y vertical de 72 ppp.

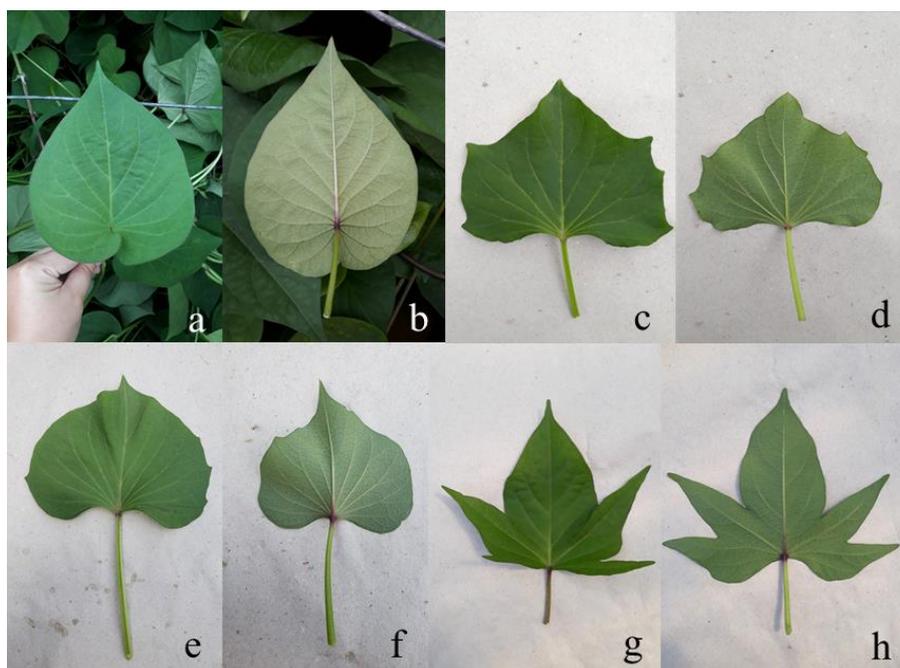


Fig. 1. Haz y envés de hojas de cuatro genotipos de batata, respectivamente: Arapey INIA (a y b); Beauregard (c y d); Boni INTA (e y f) y Morada INTA (g y h)

Este conjunto de datos posee Licencia Creative Commons Atribución 4.0 Internacional [17], según la cual, será necesario reconocer la autoría de las imágenes de forma obligatoria en cualquier explotación que se haga de las mismas. Para el uso del conjunto de datos es necesario realizar un pedido del set de datos a vilanova.antonella@inta.gob.ar.

Es intención continuar desarrollando esta fuente de imágenes, agregando nuevas fotos de los cultivares incluidos en este trabajo como así también incorporando nuevos cultivares al set y ajustando las variables relacionadas a la toma de las mismas.

Bibliografía citada

- [1] G. Loebenstein, G. Thottappilly, S. Fuentes, and J. Cohen, "Virus and phytoplasma diseases," in *The Sweetpotato*, Springer Netherlands, 2009, pp. 105–134.
- [2] D. F. Austin, "The taxonomy, evolution and genetic diversity of sweet potatoes and related wild species," *Explor. maintenance, Util. sweetpotato Genet. Resour.*, pp. 27–60, 1988.
- [3] H. Martí, *Problemas y oportunidades para el cultivo de batata ante la intensificación de algunos planteos productivos y la incorporación de cultivares más susceptibles a enfermedades. En: Sanidad en cultivos intensivos 2013*, E.E.A. San Pedro. 2013.
- [4] FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO), "FAOSTAT 2019." [Online]. Available: <http://www.fao.org/faostat/es/#data/QC>. [Accessed: 21-Aug-2020].
- [5] R. Bourke and V. Vlassak, *Estimates of Food Crop Production in Papua New Guinea*. 2004.
- [6] C. A. Clark *et al.*, "Sweetpotato viruses: 15 years of progress on understanding and managing complex diseases," *Plant Dis.*, vol. 96, no. 2, pp. 168–185, Feb. 2012.
- [7] C. C. Teow, V.-D. Truong, R. F. McFeeters, R. L. Thompson, K. V Pecota, and G. C. Yencho, "Antioxidant activities, phenolic and β -carotene contents of sweet potato genotypes with varying flesh colours," *Food Chem.*, vol. 103, no. 3, pp. 829–838, 2007.
- [8] R. Castillo Matamoros, A. Brenes Angulo, P. Esker, and L. Gómez-Alpízar, "Evaluación agronómica de trece genotipos de camote (*Ipomoea batatas* L.)," *Agronomía Costarricense*, vol. 38. scielo, pp. 67–81, 2014.
- [9] SAGPyA, "Información de Producción Hortícola. Dirección de Mercados Agroalimentarios." 2008.
- [10] P. R. Pardina, A. Luque, C. Nome, E. L. Colomba, S. F. Delgado, and L. Di Feo, "First report of Sweet potato leaf curl virus infecting sweet potato in Argentina," *Australas. Plant Dis. Notes*, vol. 7, no. 1, pp. 157–160, 2012.
- [11] J. Berger, A. Tamada, and R. Feltan, "Aplicación móvil para la identificación de variedades de Manihot esculenta Crantz cultivadas en Misiones mediante técnicas de deep learning," 2019.
- [12] J. A. Redolfi, D. González Dondo, J. A. Pucheta, and L. R. Canali, "Clasificación de variedades de semillas de trigo usando visión por computadora," in *VIII Congreso Argentino de AgroInformática (CAI-2016)-JAIIO 45 (Tres de Febrero, 2016)*, 2016.
- [13] C. M. Delfin, C. J. González, and J. C. O. Rojas, "Clasificación de manzanas

utilizando visión artificial y redes neuronales artificiales,” *Ing. y Región*, no. 20, pp. 52–57, 2018.

- [14] J. Elizondo, *Fundamentos para el procesamiento de imágenes*. Uabc, 2005.
- [15] J. Torres, *DEEP LEARNING Introducción práctica con Keras*. 2018.
- [16] D. Image, P. Dip, and I. P. Pdi, *Procesamiento y Análisis Digital de Imágenes Tareas fundamentales de DIP*. 2011.
- [17] “Creative Commons — Attribution 4.0 International — CC BY 4.0.” [Online]. Available: <https://creativecommons.org/licenses/by/4.0/legalcode>. [Accessed: 21-Aug-2020].