

Segmentación de imágenes RGB para aplicación selectiva sobre barbecho

Baruffaldi Bruno, Juan Ignacio Cavalieri y Juan Manuel Baruffaldi

Deepagro, Rosario— www.deepagro.com.ar — baruffaldibruno@gmail.com

Abstract. Este trabajo aborda la problemática de reconocimiento de malezas sin presencia de cultivos por medio de cámaras RGB. Se plantean diversas soluciones posibles trabajando con índices de vegetación o similares, cambios de espacios de colores y algoritmos de inteligencia artificial; y se analizan los niveles de predicción de cada uno, problemas surgidos y costo computacional de las soluciones para ser implementadas en un sistema embebido de medianas prestaciones. Se busca poder trabajar en tiempo real, con cámaras instaladas en el ala del equipo pulverizador a una velocidad promedio de 20 km/h y con dispositivos de bajo costo comparados con las cámaras multiespectrales.

Keywords: detección de malezas · aprendizaje automatizado · barbecho químico.

1 Introducción

El uso de herbicidas permite al productor agropecuario un control eficiente de las malezas. Sin embargo, su empleo debería minimizarse por el impacto ambiental que esto genera. A su vez, una reducción de dichos volúmenes representaría un importante beneficio económico. En Argentina, la aplicación de herbicidas sobre barbecho se realiza en forma uniforme en la totalidad del campo, no obstante el grado de enmalezamiento no es homogéneo en todo el lote. En muchos casos hay presencia de malezas solo en el 30% de la superficie total del lote en estadio de barbecho, por lo que la aplicación selectiva de agroquímicos podría representar un enorme ahorro en el uso de estos productos durante esta etapa [1].

Durante la última década, se han desarrollado distintos sistemas de pulverización selectiva que utilizan métodos de detección basados principalmente en la información contenida en la zona del espectro infrarrojo de frecuencias que refleja el suelo y las plantas. Uno de ellos es el método RR/NIR que consigue distinguir malezas sobre el suelo [2]. La información de las distintas bandas de frecuencia se suele combinar para formar distintos índices de vegetación, los cuales permiten obtener una nueva imagen donde se destacan fácilmente los puntos relacionados con parámetros de las coberturas vegetales. Por lo general,

estos índices se pueden expresar como una fórmula matemática. Uno de los más utilizados para esta tarea es el Índice de Vegetación con Diferencia Normalizada (NDVI), desarrollado en 1975 para aprovechar los datos espectrales de imágenes satelitales para evaluar la agricultura y la silvicultura mundiales. Más allá del NDVI existen muchos otros índices de vegetación menos conocidos desarrollados para propósitos específicos. Por ejemplo, el índice de vegetación ajustado al suelo (SAVI) [3] es una variante del NDVI destinado a manejar el caso en el que una fracción significativa del suelo no está cubierta de vegetación [4].

Una forma de abordar este problema es a través de cámaras multispectrales capaces de captar cambios en las bandas de radiación infrarroja para detectar las malezas de forma precisa. Sin embargo, en la actualidad estos tipos de sensores tienen asociado un costo muy elevado. A menudo, la misma información se puede obtener a partir de cámaras RGB de menor costo, aunque ésta es más sutil y difícil de identificar dado que no se conoce un índice de vegetación robusto y confiable que se base únicamente en los canales RGB visibles, y funcione en ambientes no controlados como es el campo.

Otra metodología, tal vez menos utilizada, pero que parte de una imagen RGB común, es la de encontrar un espacio de color óptimo que permita diferenciar el color verde en una imagen [8]. Sin embargo, al igual que ocurre con los índices, las condiciones ambientales provocan apariciones de sombras, píxeles sobresaturados, bajos contraste, entre otras. Estos factores pueden ocasionar cambios en los colores captados, lo cual imposibilita fijar un umbral para la segmentación sin correr el riesgo de cometer errores graves en determinadas situaciones.

El propósito de esta investigación es, entonces, estudiar diferentes métodos de segmentación que partan de información obtenida a través de cámaras RGB. Se busca que la solución sea lo suficientemente robusta para funcionar en condiciones no controladas de trabajo (al aire libre, sin internet y a diversas horas del día) y que permita una implementación que posibilite la aplicación selectiva de herbicidas sobre barbechos químicos a una velocidad de trabajo de 20 km/h.

2 Metodologías

2.1 Obtención de imágenes y etiquetado

Para este estudio se utilizaron un conjunto de imágenes digitales que poseen una resolución de 1280 x 720 píxeles y fueron tomadas en campos situados al sur de la provincia de Santa Fe en los cuales no había ningún cultivo presente.

Estas imágenes fueron adquiridas utilizando una cámara e-CAM30_CUNANO RGB de 3.4MP que se encontraba instalada en el ala de un equipo pulverizador. La misma estaba situada a 1.5 metros de altura y con un ángulo de 50 grados

respecto al suelo mientras se desplazaba a velocidades de entre 5 y 20 kilómetros por hora.

Las imágenes fueron tomadas durante los meses de Abril y Mayo de 2020 a la luz del día, bajo condiciones no controladas, en algunos casos recibiendo la luz directa del sol y en otros casos se tomaron en días nublados. Esto con la intención de diversificar las condiciones de iluminación, porcentaje de cobertura de verde, y la presencia de sombras en las imágenes.

Para este trabajo se seleccionaron 200 imágenes tratando de variar las condiciones de luminosidad mencionadas anteriormente y fueron etiquetadas manualmente bajo la categorías de “malezas” y “suelo”. Este proceso de etiquetado se basó en la selección de regiones interés, donde cada región representa un área cuadrada de 25 píxeles que pertenecen a la misma categoría. En total, se obtuvo un conjunto de entrenamiento de aproximadamente 90000 píxeles pertenecientes a la categoría “malezas” y 90000 píxeles que corresponden a la categoría “suelo”.

2.2 Procesamiento

Se analizaron distintos índices, que emulan el NDVI a través de los canales RGB, y espacios de colores para clasificar los datos. Sin embargo, se decidió priorizar los índices ya que los pasajes a distintos espacios de colores suelen requerir un mayor tiempo de cómputo.

En particular se analizaron el índice de verdor triangular (TGI)[4], el índice de resistencia atmosférica visible (VARI)[4], el índice de exceso de verde con exceso de rojo (ExG–ExR)[5] y el índice de diferencia normalizada verde-rojo (NGRDI)[6]. También se tuvieron en cuenta 8 espacios de colores (RGB normalizado, HLS, HSV, Lab, Luv, XYZ, YCrCb, Yuv)[7]. Se tomó parte de los datos etiquetados previamente y se generaron histogramas a partir de las diferentes variables para analizar mejor la información que estas contienen.

En muchos casos, es posible utilizar un valor de corte para separar los píxeles en malezas y suelo en base a los valores de estos índices. Sin embargo, este umbral puede variar drásticamente de una imagen a otra dependiendo de las condiciones de luz y presencia de verde por lo que es imposible utilizar un valor precalculado (ver Fig. 1 y Fig. 2) [9].

Una primera aproximación para evitar utilizar valores de umbrales fijos es emplear métodos estadísticos (Otsu [10], Método del triángulo [11] y adaptativo [12]) para encontrar dicho umbral de manera automatizada en base a la imagen que está siendo procesada en ese momento [13]. Sin embargo, estos métodos tienen ciertas restricciones implícitas (como garantizar la presencia de maleza en la imagen) que no siempre son posibles de asegurar en un ambiente no controlado (ver Fig. 3).

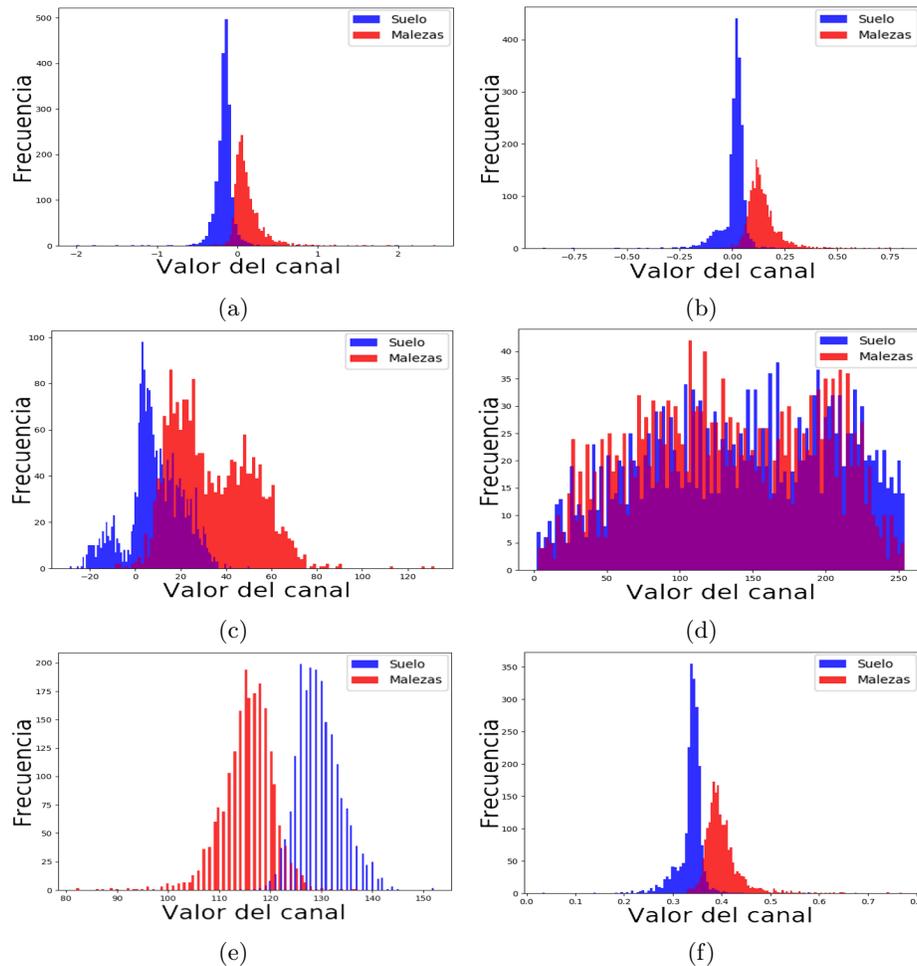


Fig. 1: Histogramas de frecuencia de pixeles pertenecientes a la clase suelo(azul) y maleza(rojo) de distintas variables. (a) Histograma del indice Exg-Exr. (b) Histograma del indice GLI. (c) Histograma del indice TGI. (d) Histograma del canal L del espacio Lab. (e) Histograma del canal a del espacio Lab. (f) Histograma del canal G del espacio RGB normalizado.

2.3 Experimentación

Por los motivos anteriormente mencionados se decidió utilizar modelos sencillos de machine learning que fuesen lo suficientemente potentes para lograr tasas de errores bajas y lo suficientemente simples para no afectar la performance del sistema. Los dos modelos seleccionados para esta tarea son árboles de decisión y redes neuronales.

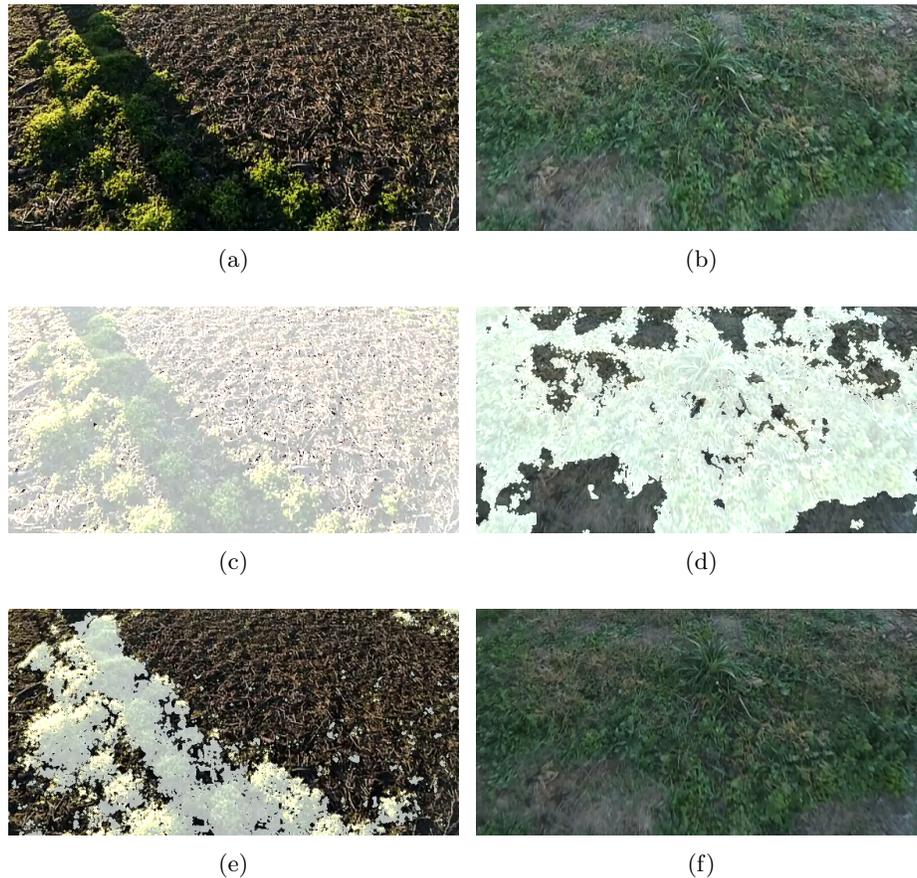


Fig. 2: Imágenes de muestra originales y con umbrales globales aplicados. (a) Imagen original con variación de iluminación. (b) Imagen original con poca iluminación. (c) Umbral en 0.01 del índice GLI aplicado a la imagen a. (d) Umbral de 0.01 del índice GLI aplicado a la imagen b. (e) Umbral de 0.54 del índice GLI aplicado sobre la imagen a. (f) Umbral de 0.54 del índice GLI aplicado sobre la imagen b.

Antes de entrenar los modelos fue necesaria la selección de los índices o canales de los distintos espacios de colores para ser utilizados como valores de entrada en los modelos. Esto es importante pues en el aprendizaje automatizado la alta dimensionalidad de un problema está asociada con el sobreajuste y la pérdida del rendimiento. En particular, para decidir cuáles iban a ser los valores de entrada para estos modelos se utilizó un método de selección de variables basado en información mutua [14]. Por cuestiones de rendimiento, este método se aplicó sobre un conjunto inicial que contenía a los 5 índices y todos los canales de 3 espacios de colores (Lab, HSV y RGB normalizado) que eran los que mejor

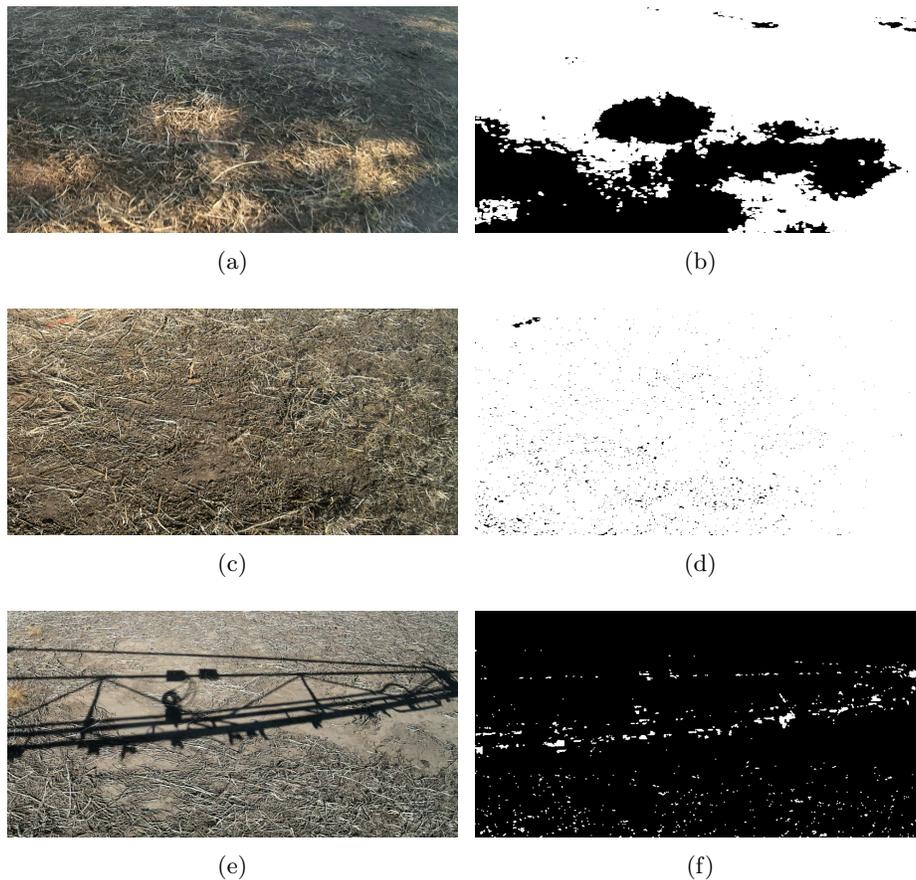


Fig. 3: Imágenes de muestra originales y con umbrales determinados por métodos estadísticos. (a) Imagen original con variación de iluminación. (b) Segmentación producida por el algoritmo Otsu sobre la imagen 'a'. (c) Imagen original sin presencia de malezas. (d) Segmentación producida por el método del triángulo sobre la imagen 'c'. (e) Imagen original con presencias de sombras. (f) Segmentación producida por el método adaptativo sobre la imagen 'e'.

funcionaron en trabajos previos [8]. Se obtuvo como resultado a un subconjunto de 3 variables que contenía al canal 'a' del espacio Lab y a los índices TGI y Exg-Exr. A dichas variables se les aplicó una normalización para que sus valores varíen dentro del rango [0, 1].

Los árboles de decisión son modelos de machine learning para aproximar funciones objetivo. Este modelo puede representarse como conjuntos de reglas condicionales y resultan muy útiles para problemas de clasificación. Estos métodos de

aprendizaje figuran entre los algoritmos de inferencia inductiva más populares y se han aplicado con éxito a una amplia gama de tareas [15].

Para esta tarea particular se entrenaron árboles de decisión a partir de los datos recolectados con los índices y canales seleccionados. Este modelo obtuvo muy buenos resultados con una profundidad de 4 niveles, lo que brinda una muy buena relación entre reconocimiento y rendimiento del sistema.

Las redes neuronales proporcionan un enfoque sólido para aproximar funciones. Para cierto tipo de problemas, como la interpretación de datos complejos de sensores del mundo real, las redes neuronales artificiales se encuentran entre los métodos de aprendizaje más robustos y eficaces [15].

En este problema se utilizó una arquitectura de red simple para minimizar el impacto del modelo en los tiempos de ejecución del sistema. En particular, se empleó una capa de entrada de 3 neuronas que recibe los valores de entrada, dos capas ocultas de 5 y 3 neuronas y una capa de salida de 1 neurona. Este modelo fue entrenado a través del descenso por el gradiente con un learning-rate de 0.001 y un momentum de 0.9. Se aplicó la función ReLU como función de activación en todas las capas, a excepción de la capa de salida en la cual se optó por la función sigmoide. Se utilizó la entropía cruzada binaria como función de error de este modelo.

3 Resultados

Dentro de los resultados obtenidos, se optó por priorizar dos características de los modelos. Por un lado se trató de minimizar la complejidad de los mismos para reducir los tiempos de ejecución del sistema, manteniendo unas tasas de errores bajas y, por otro, se dio una gran importancia al miss-rate [16] que estos presentaron. Este parámetro nos permite identificar cuáles fueron los porcentajes de falsos negativos obtenidos por cada modelo, es decir, cuántos píxeles que pertenecen a la clase “maleza” fueron clasificados como “suelo”. Esto es elemental por los fines prácticos, ya que se debe tratar de minimizar la posibilidad de que una maleza no sea identificada y por consecuencia no se le aplique herbicida.

Se contaba con un conjunto de datos de 180000 píxeles etiquetados, de los cuales se decidió apartar 33320 píxeles seleccionados de manera aleatoria manteniendo un balance entre las clases para validar el funcionamiento de los modelos. El resto de los datos fue utilizado como un conjunto de entrenamiento.

En la Tabla 1 se pueden ver los resultados obtenidos a partir de los modelos seleccionados. Se debe tener en cuenta que como los datos fueron etiquetados manualmente, estos pueden estar sujetos a un cierto nivel de ruido que es introducido por errores en el proceso de etiquetado. En las Tablas 2 y 3 se muestran las matrices de confusión de los modelos construidas a partir del grupo de píxeles de validación.

Modelos	Accuracy	Miss-Rate	F1
Árboles de decisión	94.50%	3.00%	94.65%
Redes neuronales	94.09%	5.00%	94.29%

Tabla 1: Rendimiento de los modelos utilizados.

Árboles de decisión		Predicción	
		Suelo	Maleza
Clase Real	Suelo	15300	1360
	Maleza	468	16192

Tabla 2: Matriz de confusión del árbol de decisión.

Para medir los tiempos de procesamiento se ejecutaron los modelos en un dispositivo Jetson Nano de Nvidia, que cuenta con un procesador ARM Cortex-A57 de 4 núcleos de 1.43 GHz, una placa de video Maxwell de 128 núcleos de 921 MHz y un peso de 141 gramos [17].

Para realizar una evaluación de los tiempos de procesamiento obtenidos se deben tener en cuenta las condiciones externas al software, que forman parte del sistema final. Estas condiciones nos permiten obtener una tasa de frames mínima, necesaria para garantizar que el equipo es capaz de sensor cada porción de suelo, al menos una vez.

Por un lado, la cámara utilizada tiene un campo de visión (FOV) [18] vertical de 70 grados y se encuentra a una altura de 1.5 metros con un ángulo de 40 grados con respecto al suelo. Sin embargo, la región de suelo que se analiza en cada frame abarca 1.38 metros, desde el borde inferior del frame hasta la mitad del mismo. La parte superior se descarta debido a que la distancia a la que se encuentran los objetos en esa zona de la imagen dificulta la aplicación de agroquímico de forma precisa. Por otro lado, la velocidad de trabajo del equipo pulverizador

Redes Neuronales		Predicción	
		Suelo	Maleza
Clase Real	Suelo	15647	1013
	Maleza	895	15765

Tabla 3: Matriz de confusión de la red neuronal.

que se pretende alcanzar es de 20 km/h, lo que equivale a 5.55 m/s. Todo esto implica que, para no dejar ninguna porción de suelo sin analizar, el tiempo de procesamiento de cada frame debe ser inferior a 248 milisegundos.

En la práctica se midieron los tiempos de los modelos sobre 1000 imágenes de una resolución de 1280x720, procesando únicamente la región inferior de cada frame. En promedio, se necesitan 29.12 milisegundos para segmentar cada imagen con los árboles de decisión, lo que brinda una tasa de 34 frames por segundo, y 59.37 milisegundos con las redes neuronales, alcanzando una tasa de 16 frames por segundo. Para procesar cada frame en su totalidad, se necesitan 74.77 milisegundos con los árboles de decisión y 201.14 milisegundos con las redes neuronales.

4 Conclusiones

Abordar problemas de procesamiento de imágenes con técnicas de segmentación tradicionales pueden generar buenos resultados en ambientes controlados, sin embargo, las condiciones que se presentan en ambientes externos, como el campo, dificultan mucho más su resolución.

Los modelos de machine learning son lo suficientemente robustos para brindar buenos resultados en estas condiciones. Es interesante analizar la combinación de información que necesitan tomar los modelos y con qué arquitectura se deben diseñar para poder generar soluciones que funcionen en tiempo real con un poder de cómputo acotado. Las técnicas de segmentación de imágenes pueden ser muy útiles si se combinan con estos modelos (ver Fig. 4).

Analizando los tiempos de ejecución se puede concluir que el sistema es lo suficientemente eficiente para funcionar en equipos pulverizadores cuya velocidad de trabajo alcance los 20 km/h.

5 Trabajos a futuro

Como trabajo a futuro sería interesante investigar la aplicación de técnicas de Data Augmentation para generar mayor robustez en los datos de entrenamiento y PCA para realizar una reducción de la dimensionalidad del problema abordado mediante la extracción de características en lugar de la selección de variables. Estos métodos de extracción de características se basan en la generación de nuevas variables y no solamente a la selección de variables ya existentes a partir de un conjunto [19].

En segunda instancia sería interesante investigar la utilización de técnicas más avanzadas de deep learning para abordar este problema [20].



(a) Original



(b) Segmentada



(c) Original



(d) Segmentada



(e) Original



(f) Segmentada



(g) Original



(h) Segmentada



(i) Original



(j) Segmentada

Fig. 4: Segmentación mediante el modelo de árbol de decisión.

Referencias

1. Moltoni, Andrés Fernando, and Luciana Andrea Moltoni. "Pulverización selectiva de herbicidas. implicancias tecnológicas y económicas de su implementación en la Argentina." *Curso Internacional de Agricultura de Precisión. 6. Expo de Máquinas Precisas. 1. 2006 07 25-27, 25 al 27 de julio de 2006. Manfredi, Córdoba. AR. (2006).*
2. Moltoni, Andrés Fernando, and Gerardo Masiá. "Site-Specific Weed Control: Daily Performance of a Weed Detector System Designed by Researchers of INTA." *CIGR-International Conference of Agricultural Engineering-XXXVII Congresso Brasileiro de Engenharia Agrícola. CONBEA. 2008.*
3. Huete, Alfredo. "Huete, AR A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment.*" *Remote sensing of environment* 25 (1988): 295-309.
4. McKinnon, Tom, and Paul Hoff. "Comparing RGB-based vegetation indices with NDVI for drone based agricultural sensing." *Agribotix. Com* (2017): 1-8.
5. Meyer, George E., and João Camargo Neto. "Verification of color vegetation indices for automated crop imaging applications." *Computers and electronics in agriculture* 63.2 (2008): 282-293.
6. Hunt, E. Raymond, et al. "Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status." *Precision Agriculture* 6.4 (2005): 359-378.
7. García-Mateos, G., et al. "Study and comparison of color models for automatic image analysis in irrigation management applications." *Agricultural Water Management* 151 (2015): 158-166.
8. Hernández-Hernández, J. L., et al. "Optimal color space selection method for plant/soil segmentation in agriculture." *Computers and Electronics in Agriculture* 122 (2016): 124-132.
9. Guo, Wei, Uday K. Rage, and Seishi Ninomiya. "Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model." *Computers and electronics in agriculture* 96 (2013): 58-66.
10. Kurita, Takio, Nobuyuki Otsu, and N. Abdelmalek. "Maximum likelihood thresholding based on population mixture models." *Pattern recognition* 25.10 (1992): 1231-1240.
11. Zack, Gregory W., William E. Rogers, and Samuel A. Latt. "Automatic measurement of sister chromatid exchange frequency." *Journal of Histochemistry & Cytochemistry* 25.7 (1977): 741-753.
12. Lei, Jian-feng, and Wei Wang. "Research and implementation of image threshold segmentation based on OpenCV." *Mod Electron Tech* 36.24 (2013): 72-76.
13. Tajima, Ryosuke, and Yoichiro Kato. "Comparison of threshold algorithms for automatic image processing of rice roots using freeware ImageJ." *Field Crops Research* 121.3 (2011): 460-463.
14. Kwak, Nojun, and Chong-Ho Choi. "Input feature selection for classification problems." *IEEE transactions on neural networks* 13.1 (2002): 143-159.
15. Mitchell, Thomas M. "Machine learning." (1997).
16. Tharwat, Alaa. "Classification assessment methods." *Applied Computing and Informatics* (2018).
17. Nvidia. "Data sheet Nvidia Jetson Nano System-on-Module." (2020)
18. E-con Systems. "e-CAM30_CUNANO Lens Datasheet." (2019)
19. Howley, Tom, et al. "The effect of principal component analysis on machine learning accuracy with high dimensional spectral data." *International Conference on Innovative Techniques and Applications of Artificial Intelligence. Springer, London, 2005.*

20. Milioto, Andres, Philipp Lottes, and Cyrill Stachniss. "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs." 2018 IEEE international conference on robotics and automation (ICRA). IEEE, 2018.