

PREDICCIÓN DE RENDIMIENTO DE COSECHAS UTILIZANDO APRENDIZAJE PROFUNDO MULTIMODAL

Data de aceite: 02/05/2024

Luis Roberto Jácome Galarza

CiDiS - Centro de Investigación,
Desarrollo e Innovación de Sistemas
Computacionales. Escuela Superior
Politécnica del Litoral, ESPOL
Guayaquil, Ecuador
Facultad de Ciencias Técnicas,
Universidad Internacional Del Ecuador
UIDE, Quito, Ecuador

RESUMEN: La agricultura de precisión es una práctica vital para mejorar la producción de cosechas. El presente trabajo tiene como objetivo desarrollar un modelo multimodal de aprendizaje profundo que es capaz de producir un mapa de salud de cosechas. El modelo recibe como entradas imágenes multi-espectrales y datos de sensores de campo (humedad, temperatura, estado del suelo, etc.) y crea un mapa de rendimiento de la cosecha. La utilización de datos multimodales tiene como finalidad extraer patrones ocultos del estado de salud de las cosechas y de esta manera obtener mejores resultados que los obtenidos mediante los índices de vegetación.

PALABRAS-CLAVE: Agricultura de precisión; sensores remotos; aprendizaje profundo multimodal; IoT; agentes inteligentes; computación aplicada.

CROP YIELD PREDICTION UTILIZING MULTIMODAL DEEP LEARNING

ABSTRACT: Precision agriculture is a vital practice for improving the production of crops. The present work is aimed to develop a multimodal deep learning model that is able to produce a prediction map of the health of crops. The model takes multispectral images and field sensor data (humidity, temperature, soil status, etc.) as an input and creates a yield map of a crop. The utilization of multimodal data is aimed to extract hidden patterns in the status of crops and in this way obtain better results than the use of vegetation indices.

KEYWORDS: Precision agriculture; remote sensing; convolutional neural networks; recurrent neural networks; multimodal deep learning; IoT; intelligent agents; applied computation.

EL PROBLEMA Y LA RAZÓN DE SU IMPORTANCIA O RELEVANCIA

La predicción de rendimiento de cosechas es extremadamente complicada debido a su dependencia de múltiples factores como el genotipo de las cosechas, factores ambientales, prácticas de manejo de cultivos, variedades de plantas, etc. [1].

En el proceso de crecimiento de las plantas, los nutrientes minerales juegan un papel importante. Los macronutrientes incluyen calcio, potasio, nitrógeno, magnesio, azufre y fósforo; mientras que los micronutrientes incluyen boro, hierro, manganeso, cloruro, zinc y molibdeno. La deficiencia de estos nutrientes afectan el crecimiento, rendimiento y calidad de las plantas y cosechas [2].

La estimación precisa del rendimiento de las cosechas y el manejo optimizado de nitrógeno (fertilizantes) es esencial en la agricultura [3], estos dos aspectos pueden ayudar a asegurar la producción de alimentos y la optimización del uso de fertilizantes lo que es a su vez beneficios para el ambiente.

El presente trabajo propone el diseño e implementación de un modelo de aprendizaje profundo multimodal que permita la predicción de rendimiento de cosechas utilizando datos espacio-temporales de sensores de campo e imágenes multi-espectrales, de esta forma, se quiere mejorar la productividad a través del conocimiento del estado de salud de cultivos que en muchos casos es difícil de predecir mediante la observación humana.

Un modelo más preciso de predicción de cultivos conlleva implicaciones muy significativas en la agricultura de precisión, puesto que los agricultores tendrían la posibilidad de detectar falencias de los cultivos en etapas tempranas de la cosecha, lo que permite tomar acciones correctivas; también es relevante la optimización de recursos como agua, fertilizantes, pesticidas, etc.; finalmente permite optimizar la gestión de los inventarios de cosechas en actividades como planificación, uso de bodegas, estimación de precios, etc.

LOS OBJETIVOS Y TÓPICOS DE INVESTIGACIÓN ASOCIADOS

Objetivo general

Diseñar e implementar un modelo de aprendizaje profundo multimodal para predicción de rendimiento de cosechas utilizando imágenes multi-espectrales y datos de sensores de campo.

Objetivos específicos

- Evaluar arquitecturas de aprendizaje profundo para predicción de series de tiempo y procesamiento de imágenes.
- Diseñar e implementar un modelo de aprendizaje profundo multimodal con imágenes multi-espectrales y datos de sensores de campo.
- Validar el modelo de aprendizaje profundo multimodal y compararlo con otros enfoques que constituyen el estado del arte.

ESTADO DEL ARTE

En lo referente al estado del arte, se tiene los siguientes trabajos relacionados a predicción de rendimiento de cultivos:

En el trabajo en [1], se presenta un enfoque de aprendizaje profundo para predicción de cultivos de maíz y soya a través del cinturón de maíz en los Estados Unidos, basado en datos ambientales y prácticas de manejo agrícola.

Un nuevo marco de trabajo basado en aprendizaje profundo y vehículos aéreos no tripulados es utilizado para inferir en la calidad de los campos de arroz [5].

En el trabajo en [6], utilizan una arquitectura de aprendizaje profundo para reconocer y clasificar las etapas fenológicas de varios tipos de plantas basado solamente en datos visuales.

El trabajo en [22], presenta un listado de conjuntos de datos relacionados a agricultura (TABLE I). Si bien estos conjuntos de datos son usados para predecir el rendimiento de cosechas, no contienen información de sensores de campo.

Por otro lado, el trabajo en [34] presenta una revisión literaria en el cual se destacan conjuntos de datos de aprendizaje multimodal y sus respectivas aplicaciones (TABLE II), y diversas aplicaciones de aprendizaje profundo multimodal (TABLE III).

#	Conjunto de datos relacionados con agricultura	
	Conjunto de datos (Dataset)	Descripción
1	Image-Net Dataset	Images of various plants (trees, vegetables, flowers)
2	ImageNet Large Scale Visual Recognition Challenge (ILSVRC)	Images that allow object localization and detection
3	University of Arkansas, Plants Dataset	Herbicide injury image database
4	EPFL, Plant Village Dataset	Images of various crops and their diseases
5	Leafsnap Dataset	Leaves from 185 tree species from the Northeastern United States
6	LifeCLEF Dataset	Identity, geographic distribution and uses of plants
7	PASCAL Visual Object Classes Dataset	Images of various animals (birds, cats, cows, dogs, horses, sheep etc.)
8	Africa Soil Information Service (AFSIS) dataset	Continent-wide digital soil maps for sub-Saharan Africa
9	UC Merced Land Use Dataset	A 21 class land use image dataset
10	MalayaKew Dataset	Scan-like images of leaves from 44 species classes
11	Crop/Weed Field Image Dataset	Field images, vegetation segmentation masks and crop/weed plant type annotations.
12	University of Bonn Photogrammetry, IGG	Sugar beets dataset for plant classification as well as localization and mapping
13	Flavia leaf dataset	Leaf images of 32 plants.
14	Syngenta Crop Challenge 2017	2,267 of corn hybrids in 2,122 of locations between 2008 and 2016, together with weather and soil conditions

TABLE I. Conjuntos de datos relacionados con agricultura

#	Conjuntos de datos de aprendizaje multimodal		
	Conjunto de datos (Dataset)	Descripción	Aplicación
1	UTD-MHAD	Depth and inertial sensor data	Human action recognition
2	ChaLearn looking at people	RGB-D, audio, skeletal pose	Human action recognition
3	Berkeley MHAD	Multiviewpoint RGB-D and skeletal pose data	Human action recognition
4	MHRI data set	Chest, top RGB-D, face, video, and audio	Human-robot interaction
5	H-MOG	Nine smartphone sensors and interaction data	Continuous authentication in smartphones
6	RECOLA	Audio, visual, and physiological	Emotion recognition
7	MHEALTH	Accelerometer, electrocardiogram, magnetometer, and gyroscopes	Health monitoring
8	Pinterest Multimodal	Images and text (40M)	Multimodal word embeddings
9	MM-IMDb	Video, images, and text metadata	Movie genre prediction
10	FCVID	Video and audio	Action recognition
11	KITTI	Stereo gray- and color video, 3-D-LIDAR, inertial and GPS navigation data	Autonomous driving
12	KinectFaceDB	RGB-D and facial landmarks	Face recognition
13	Oxford RobotCar	Six cameras, LIDAR, GPS, and inertial navigation data	Autonomous driving
14	Multimodal BRATS	T2-, FLAIR-, post-Gadolinium T1-MRI, perfusion, and diffusion MRI and MRSI	Brain tumor segmentation

TABLE II. Conjuntos de datos de aprendizaje multimodal y sus aplicaciones

Los modelos y conjuntos de datos mencionados en TABLE II y TABLE III no son relacionados a la agricultura. En contraste, el conjunto de datos llamado Ladybird Cobbity 2017 Brassica [35] sí está relacionado a la agricultura (plantas de coliflor y brócoli) y contiene imágenes RGB, imágenes hiper-espectrales, datos manuales (valores SPAD y longitud de plantas) y datos de sensores de campo (temperatura, humedad, condiciones del suelo, etc.).

Modalidades	Aplicaciones de aprendizaje profundo multimodal		
	<i>Problema</i>	<i>Método de fusión</i>	<i>Arquitectura</i>
Audio, video	Clasificación de voz	Intermedia	Sparse RBM
Imagen, texto	Anotación de imágenes	Intermedia	DBN
Imágenes médicas, descripción textual	Recuperación de imágenes médicas basado en contenido	Intermedia	DBM
Expresión de genes, información de ADN y respuesta a fármacos	Agrupación de subtipos de cáncer	Intermedia	DBM
Imágenes multi-espectrales	Segmentación semántica	Temprana	FCNN
Imágenes y flujo óptico (optical flow)	Reconocimiento de acciones	Tardía	CNN
Video, audio	Reconocimiento de emociones	Tardía	CNN, RNN, SVM, AE
MRI, PET	Diagnóstico médico	Intermedia	Stacked AE, SVM
Video, audio, texto	Análisis de sentimientos	Intermedia, tardía	CNN, SVM
Intensidad, profundidad de video	Robotic grasping	Intermedia	Stacked AE, MLP
Características de video, CoordenadasGPS, dinámica de vehículo	Anticipación de actividades de conductores	Intermedia	LSTM

TABLE III. Aplicaciones de aprendizaje profundo multimodal

METODOLOGÍA PROPUESTA PARA LA RESOLUCIÓN DEL PROBLEMA

El presente proyecto se trata de un modelo de aprendizaje de máquinas que integra datos de sensores de campo e imágenes multi-espectrales. Para ello, es necesario diseñar e implementar una arquitectura multimodal de aprendizaje profundo. Por una parte, el modelo está formado por una red neuronal convolucional para procesar las imágenes multi-espectrales. Por otra parte, el modelo está conformado por una red neuronal profunda para analizar los datos de los sensores de campo; esta parte del modelo extrae los patrones temporales y no temporales de los datos de los sensores de campo. Los datos de sensores a considerar son humedad, temperatura, humedad de suelo, etc., los mismos que son usados comúnmente en agricultura de precisión.

Para realizar la experimentación se ha considerado el conjunto de datos Ladybird Cobbity 2017 Brassica [35], que contiene información de cultivos de coliflor y brócoli. De forma manual se ha obtenido el peso, altura, diámetro y el contenido relativo de agua de los cultivos; así mismo el conjunto de datos tiene las mediciones de SPAD (Soil-Plant Analysis Development) que constituye una métrica de la cantidad de nitrógeno de las plantas y que en el presente proyecto se utiliza como “ground truth” del experimento. El conjunto de datos Ladybird Cobbity 2017 Brassica contiene además datos de sensores de campo como sensores del suelo que miden conductividad eléctrica, temperatura, contenido de agua, etc.; datos temporales de una estación meteorológica con información sobre la temperatura

ambiente, velocidad y dirección del viento, humedad, presión barométrica, cantidad de lluvia, etc. Por último, el conjunto de datos mencionado presenta información de imágenes multi-espectrales de alta resolución. Se indica que las imágenes fueron adquiridas mediante el vehículo terrestre no tripulado denominado “Ladybird” que es una plataforma robótica para agricultura. Las imágenes multi-espectrales fueron capturas mediante una cámara Resonon Pika XC2. El conjunto de datos, además contiene imágenes estereoscópicas tomadas con 2 cámaras Grasshopper3 12MP GS3 - U3 - 120S6C – C, e imágenes térmicas obtenidas con una cámara Xenics Gobi – 640. Las imágenes estereoscópicas y térmicas no son consideradas en el experimento puesto que resultan redundantes con la información de las imágenes multi-espectrales.

Por su parte, el modelo propuesto de aprendizaje profundo está conformado por 2 modelos de redes neuronales profundas. Por una parte, el primer modelo contiene información de sensores ubicados en ciertas locaciones de un campo agrícola. La información de los sensores y su ubicación aportan conocimiento a los mapas de rendimiento de los cultivos generados por el segundo modelo que es alimentado con imágenes multi-espectrales.

Los datos de sensores de campo son tratados con un modelo basado en la arquitectura GeThR/Net [36]. Este modelo, primeramente, extrae los patrones temporales de cada modalidad de sensor utilizando redes neuronales recurrentes de tipo LSTM (Long Short-Term Memory). El modelo también extrae patrones no temporales de cada una de las modalidades de sensores de campo, para ello utiliza la arquitectura de MLP (Multi Layer Perceptron). Finalmente, este primer modelo combina los patrones temporales y no temporales para dar una métrica final de predicción (Fig. 1).

Por su parte, las imágenes multi-espectrales son tratadas por un modelo basado en la arquitectura U-GAN [37]. Este modelo de tipo GAN (Generative Adversarial Network) está conformado por un generador y un discriminador (Fig. 2). El generador es entrenado con imágenes multi-espectrales de la cosecha con los canales RGB (Red, Green Blue) y NIR (Near Infra Red) y las correspondientes imágenes del rendimiento de esa cosecha. El objetivo del generador es aprender la distribución de rendimiento de la cosecha y luego poder generar esos mapas de rendimiento. Por su parte, el discriminador tiene el objetivo de identificar imágenes reales y falsas (generadas) del rendimiento de cosechas.

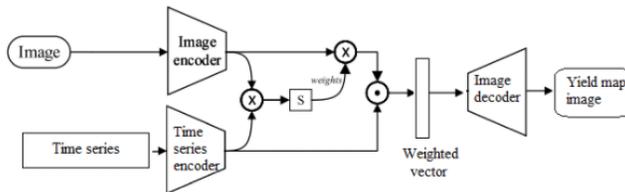


Figure 1. Parte del modelo de aprendizaje profundo propuesto que obtiene patrones temporales y no temporales de los sensores de campo por medio de una arquitectura basada en GeThR/Net [36].

El generador utiliza la arquitectura U-NET con un bloque denso (dense block) el cual está formado por un codificador (encoder) y un decodificador (decoder) con una compuerta de atención (attention gate). El discriminador utiliza un bloque denso (dense block).

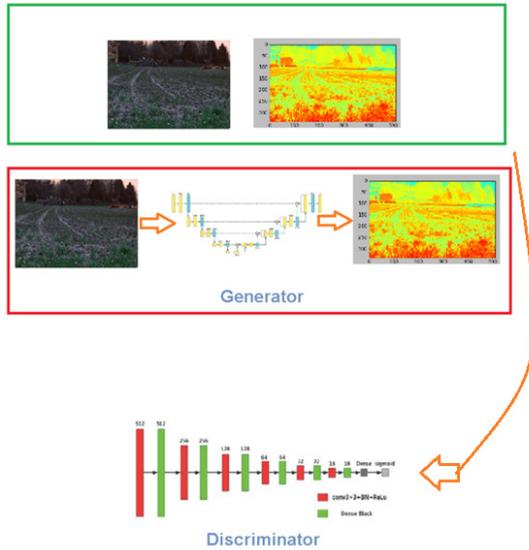


Figure 2. Parte del modelo de aprendizaje profundo propuesto que genera imágenes de rendimiento de cosechas utilizando una arquitectura basada en U-NET [37].

Finalmente, se combina los resultados de predicción de los 2 modelos mencionados anteriormente para obtener el mapa final de rendimiento de la cosecha. En la Fig. 3 se ilustra el modelo multimodal de aprendizaje profundo completo.

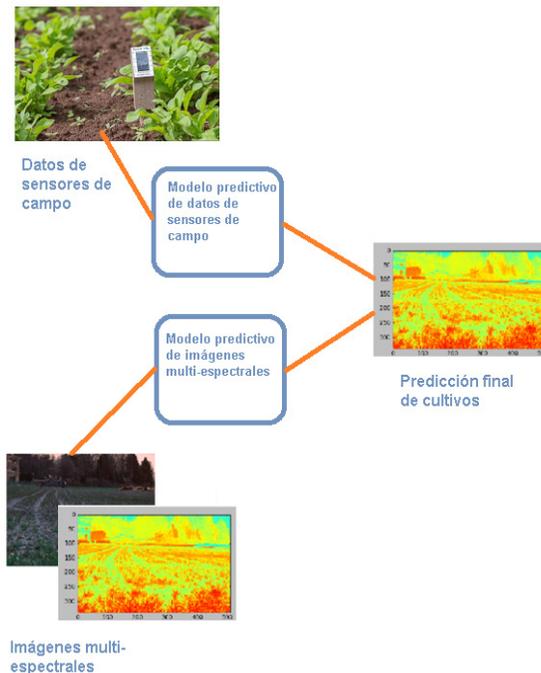


Figure 3. Modelo de aprendizaje profundo completo. Incluye el modelo de patrones temporales y no temporales de los datos de sensores de campo y el modelo generativo de imágenes multi-espectrales

RESULTADOS ESPERADOS

La expectativa del presente proyecto es que el modelo de aprendizaje profundo sea capaz de predecir el rendimiento de cosechas considerando diferentes fuentes de datos, por una parte datos de sensores de campo de una cosecha (se quiere extraer información relevante de variables obtenidas de las condiciones del ambiente que influyen directamente en el estado de salud de las cosechas) y por otro lado imágenes multi-espectrales de la misma cosecha. Se espera que el modelo de aprendizaje profundo sea capaz de superar enfoques de agricultura de precisión (por ejemplo los índices de vegetación como es el NDVI – Normalized Difference Vegetation Index) que constituyen el estado del arte en lo referente a la predicción de rendimiento de cosechas.

REFERÊNCIAS

- [1] S. Khaki, L. Wang, L., and S. Archontoulis, S., "A CNN-RNN framework for crop yield prediction," arXiv preprint arXiv:1911.09045, 2019.
- [2] T. Tran, J. Choi, T. Le, and J. Kim, "A comparative study of deep CNN in forecasting and classifying the macronutrient deficiencies on development of tomato plant," *Applied Sciences*, 9(8), 1601, 2019.
- [3] A. Chlingaryan, S. Sukkariéh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Computers and electronics in agriculture*, 151, 2018, pp. 61-69.
- [4] B. Basso, D. Cammarano, and E. Carfagna, "Review of crop yield forecasting methods and early warning systems," in *Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics*, FAO Headquarters, Rome, Italy, 2013, pp. 18-19.
- [5] N. Tri, T. Van Hoai, H. Duong, "A novel framework based on deep learning and unmanned aerial vehicles to assess the quality of rice fields," in : *International Conference on Advances in Information and Communication Technology*. Springer, Cham, 2016, pp. 84-93.
- [6] H. Yalcin, "Plant phenology recognition using deep learning: Deep-Pheno," in : *2017 6th International Conference on Agro-Geoinformatics*. IEEE, 2017, pp. 1-5.
- [7] X. Zhang, Y. QIAO, Yue, "Identification of maize leaf diseases using improved deep convolutional neural networks," *IEEE Access*, vol. 6, 2018, pp. 30370-30377.
- [8] A. Córdova, M. Barreno, J. Jácome "Precise weed and maize classification through convolutional neuronal networks," in : *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, 2017, pp. 1-6.
- [9] P. Lin, X. Li, Y. Chen, "A deep convolutional neural network architecture for boosting image discrimination accuracy of rice species," *Food and bioprocess technology*, vol. 11, no 4, 2018, pp. 765-773.
- [10] P. Lin, Y. Chen, H. Jianqiang, "Determination of the varieties of rice kernels based on machine vision and deep learning technology," in : *2017 10th International Symposium on Computational Intelligence and Design (ISCID)*. IEEE, 2017, pp. 169-172.

- [11] Y. Li, M. Qian, P. Liu, "The recognition of rice images by UAV based on capsule network," *Cluster Computing*, 2018, pp. 1-10.
- [12] R. Atole, D. Park, "A multiclass deep convolutional neural network classifier for detection of common rice plant anomalies," *International Journal of Advanced Computer Science and Applications*, vol. 9, no 1, 2018, pp. 67-70.
- [13] A. Singh, B. Ganapathysubramanian A. Singh, "Machine learning for high-throughput stress phenotyping in plants," *Trends in plant science*, vol. 21, no 2, 2016, pp. 110-124.
- [14] N. Kussul, M. Lavreniuk, S. Skakun, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no 5, 2017, pp. 778-782.
- [15] D. Hughes, M. Salathé, M., "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, 2015.
- [16] Y. Zheng, J. Kong, X. Jin, X. Wang, T. Su, and M. Zuo, "CropDeep: the crop vision dataset for deep-learning-based classification and detection in precision agriculture. *Sensors*, 19(5), 1058, 2019.
- [17] M. Nilsback, A. Zisserman, "A visual vocabulary for flower classification," *Computer Vision and Pattern Recognition*. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY, USA, June 2006, pp. 1447–1454.
- [18] N. Kumar, P. Belhumeur, A. Biswas, A. D. Jacobs, W. Kress, I. Lopez, J. Soares, "Leafsnap: A computer vision system for automatic plant species identification," in *Proceedings of the European Conference on Computer Vision*, Florence, Italy, October 2012, pp. 502–516.
- [19] J. Wegner, S. Branson, D. Hall, D. K. Schindler, P. Perona, "Cataloging public objects using aerial and street-level images-urban trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas Valley, NV, USA, 26 June–1 July 2016, pp. 6014–6023
- [20] The University of Georgia-WSFNR; The University of Georgia-CAES; USDA, "Identification Technology Program 'IPM Images,'" [Online]. Available: <https://www.ipmimages.org/>, 2017.
- [21] D. Moshou, C. Bravo, S. Wahlen, J. West, A. McCartney, J. De Baerdemaeker, H. Ramon, "Simultaneous identification of plant stresses and diseases in arable crops using proximal optical sensing and self-organising maps," *Precis. Agric.*7, 2006, pp. 149–164.
- [22] A. Kamilaris, and F. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, 147, 2018, pp 70-90.
- [23] J. Shanahan, J. Schepers, D. Francis, G. Varvel, W. Wilhelm, J. Tringe, and D. Major, "Use of remote-sensing imagery to estimate corn grain yield," *Agronomy Journal*, 93(3), 2001, pp. 583-589.
- [24] A. Vina, A. Gitelson, D. Rundquist, G. Keydan, B. Leavitt, and J. Schepers, "Monitoring maize (*Zea mays* L.) phenology with remote sensing," *Agronomy Journal*, 96(4), 2004, pp. 1139-1147.
- [25] B. Peng, K. Guan, M. Pan, and Y. Li, "Benefits of seasonal climate prediction and satellite data for forecasting US maize yield," *Geophysical Research Letters*, 45(18), 2018, pp. 9662-9671.

- [26] W. Chivasa, O. Mutanga, and C. Biradar, "Application of remote sensing in estimating maize grain yield in heterogeneous African agricultural landscapes: a review," *International journal of remote sensing*, 38(23), 2017, pp. 6816-6845.
- [27] B. Basso, D. Cammarano, and E. Carfagna, "Review of crop yield forecasting methods and early warning systems," in *Proceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics*, FAO Headquarters, Rome, Italy, July 2013, pp. 18-19.
- [28] C. Teoh, D. Hassan, M. Radzali, and J. Jafni, "Prediction of SPAD chlorophyll meter readings using remote sensing technique," *J. Trop. Agric. Food Sci*, 40(1), 2012, pp. 127-136.
- [29] <http://www.iniap.gob.ec/pruebav3/wp-content/uploads/2018/03/281-iniap-OK-baja.pdf>
- [30] <http://sipa.agricultura.gob.ec/>
- [31] <http://sipa.agricultura.gob.ec/index.php/maiz>
- [32] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Ng, "Multimodal deep learning," in *ICML*, January 2011.
- [33] N. Srivastava, and R. Salakhutdinov, R. "Multimodal learning with deep boltzmann machines," *The Journal of Machine Learning Research*, 15(1), 2014, pp. 2949-2980.
- [34] D. Ramachandram, and G. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Processing Magazine*, 34(6), 2017, pp. 96-108.
- [35] A. Bender, B. Whelan, and S. Sukkarieh, *Ladybird Cobbitty 2017 Brassica dataset*, 2019.
- [36] Gandhi, Ankit, "Gethr-net: A generalized temporally hybrid recurrent neural network for multimodal information fusion," *European conference on computer vision*. Springer, Cham, 2016.
- [37] C. Wu, Y. Zou, and Z. Yang, "U-GAN: Generative Adversarial Networks with U-Net for Retinal Vessel Segmentation," in *2019 14th International Conference on Computer Science & Education (ICCSE) IEEE*, August 2019, pp. 642-646.