Deep learning in plant science: A mini-review
Michael Alabboud 1,2*

1, Department of Horticulture, Al-Baath University, Homs, Syria
2, Department of Horticultural Sciences, UTCAN, University of Tehran, Iran

Abstract

The role that deep learning plays in modern life is undeniably essential. It is also certain that deep learning, with its various approaches, is contributing significantly to plant science. Whether by explaining the acquired data or converting and refining these data to a more profound level, deep learning techniques are pushing the frontiers of plant research further than ever before. This study is an attempt to shed light on recent advances and applications of deep learning in plant science. These applications were systematically reviewed at omics, micro/macroscopic, and population levels. Future aspects were also discussed to some extent.

Keywords: Convolutional neural networks, Deep learning applications, Plant omics

1. Introduction

Artificial intelligence (AI) is defined as the act of thinking and learning when performed by a computer program or machine. Due to its capability of simulating human intelligence and thus performing complicated human tasks, AI is steadily replacing human forces in various sectors. In science, AI is creating new frontiers in environmental [1], medical [2], pharmaceutical [3], biological [4], agricultural [5], and engineering [6] sciences.

Various methodologies are used in AI research, and classifying these methods is tricky since many of them are interconnected. However, the classification agreed upon by many scholars subdivides AI techniques into four divisions: Machine learning (ML), Natural language processing (NLP), Computer vision, and Robotics (Fig. 1). ML is a major branch of AI that gained increasing importance in the big data era [7]. Deep learning (DL) is a subset field of ML (Fig. 1). The recent rapid growth of DL is generated from the need for a real-time image processing technique, which is the main application of DL through deep and convolutional neural networks (CNN).

Figure 1. Artificial intelligence subsets.
Deep learning models are developed to tackle real-life problems in various scientific fields. The aim of this study was to review recent literature of DL applications in plant science. Furthermore, the major difficulties and future aspects of DL in plant science will be discussed.

2. Deep learning definition

Deep learning is a comprehensive tool of ML that is constituted upon the principles of artificial neural networks (ANN) with great capability to discover complex features in multilayered data. In fact, the main advantage of DL technique compared to conventional ANNs is its efficiency in automatically extracting the significant features of the analyzed data. This automation allows the researcher to focus on model architecture and results reasoning rather than spending long hours on manual feature extraction [8]. This advantage rendered DL as an ideal tool to tackle many scientific problems in health care [9], bioinformatics [10], finance [11], and agriculture [12].

3. Deep learning in plant science

DL potentials in plant science research are undeniable. The reason behind this certainty is that plant science, similar to all modern sciences, dramatically benefits from visual data and data visualization [13]. Technically, any form of visual input can be used for this technique, whether it is any form of spectral imaging or visualization of data to be interpreted as an image such as metabolic profiles and nucleic acid sequence. However, different DL techniques are suitable for the different data inputs. For instance, CNN models are ideal for high dimensional data such as spectral images inputs, while recurrent neural network (RNN) and long-/short-term memory (LSTM) are preferable for sequential data such as DNA and RNA sequences.

Many approaches can be used to tackle the subject of deep learning applications in plant science. For instance, DL applications can be classified technically based on the DL technique or input data format. However, the current review addresses the subject by classifying DL applications into three distinctive levels: Omics level, Micro/Macroscopic level, and population level.

3.1. Omics level

Oomics are a collective field of biological studies that end with the suffix -omics, such as genomics, transcriptomics, proteomics, and metabolomics. These studies address the quantification, structure, and functionality of biological molecules. Therefore, oomics provide valuable knowledge of plant organizational functionality [14]. Due to the large data usually provided by oomics studies methodologies, DL became a necessary inseparable tool of oomics data processing and reasoning.

Since oomics data is usually sequential, RNNs and LSTM are widely used to process these data. The primary purpose of DL in oomics studies is to locate and highlight unique features of interest in the studied data, such as detecting single nuclear polymorphisms (SNPs) (Fig. 2 A) and enhancers’ sequences (Fig. 2 B) [15] or to translate the obtained data (input) into other forms of information (output). For this purpose, molecular data is collectively or individually used to produce a general conceptualization of the plant morphology and phenotypic characteristics, which might have limitless breeding applications [16] (Fig. 2 C).

DL techniques are pushing oomics research forward in many aspects (Table 1); however, it is stated that other techniques, such as non-additive Gaussian kernel or simple arc-cosine kernel, might produce more reliable predictions based on oomics data compared to DL [17]. This observation might be attributed to the overall complicated DL fine-tuning. Therefore, more research should be conducted in the field of DL model’s optimization for oomics-based predictions.
Table 1. Recent deep learning applications in plant omics.

<table>
<thead>
<tr>
<th>Plant</th>
<th>Data</th>
<th>Model</th>
<th>Application</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize</td>
<td>GBS*</td>
<td>LSTM</td>
<td>SNP detection and Trait prediction</td>
<td>[18]</td>
</tr>
<tr>
<td>Eucalyptus</td>
<td>GBS</td>
<td>CNN</td>
<td>rSNPs** identification</td>
<td>[19]</td>
</tr>
<tr>
<td>Broad bean</td>
<td>GBS</td>
<td>CNN</td>
<td>Epigenomic modifications examining</td>
<td>[20]</td>
</tr>
<tr>
<td>Rice</td>
<td>DNA methylation and RNA-seq</td>
<td>CNN</td>
<td>Predicting transcriptional levels</td>
<td>[21]</td>
</tr>
<tr>
<td>Maize</td>
<td>DNA Sequence</td>
<td>CNN</td>
<td>SNP detection and Trait prediction</td>
<td>[22]</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>Nanopore DNA reads</td>
<td>BRNN*** with LSTM units</td>
<td>Cytosine methylations detection</td>
<td>[23]</td>
</tr>
<tr>
<td>Rice</td>
<td>Soybean</td>
<td>CNN</td>
<td>Phenotype prediction</td>
<td>[24]</td>
</tr>
<tr>
<td>Wheat</td>
<td>SNP</td>
<td>CNN</td>
<td>Phenotype prediction</td>
<td></td>
</tr>
<tr>
<td>Maize</td>
<td>SNP</td>
<td>Densely connected network</td>
<td>Trait prediction under various environments</td>
<td></td>
</tr>
<tr>
<td>Wheat</td>
<td>SNP</td>
<td>self-attention network (SAN)</td>
<td>Investigating plant-pathogen interaction (Barley x Powdery mildew)</td>
<td>[25]</td>
</tr>
<tr>
<td>Barley</td>
<td>Hyperspectral UV scans and secondary metabolites</td>
<td>self-attention network (SAN)</td>
<td>Investigating plant-pathogen interaction (Barley x Powdery mildew)</td>
<td>[26]</td>
</tr>
</tbody>
</table>

* GBS: Genotyping by sequencing GBS
** rSNPs: Regulatory SNPs
*** BRNN: Bidirectional recurrent neural networks
3.2. Micro/Macroscopic level

The main concern of microscopic studies is to investigate the plant on a cellular level, such as cellular organelles, full cell, and tissue research. On the other hand, macroscopic studies, in this context, address the characterizations and interactions of a whole or part of a plant. The data used for this type of study are predominantly images of any range of the electromagnetic spectrum.

Various data acquisition systems are being used in Micro/Macroscopic DL, such as visible light sensors [27], infrared (IR) and near-infrared (NIR) [28], ultraviolet [29], hyperspectral [30], and even X-rays [31]. Furthermore, other supportive techniques are being used to increase the depth of the acquired imagery such as cell staining in microscopy [32] and fluorescence imaging systems [33]. There are also some attempts to employ plant’s electrical signals measurements in DL research [34].

The applications of DL in this level are countless such as taxonomy and classification, disease/stress recognition and early warning systems, and physiological events tracking. (Table 2) covers some of the recent studies in the field.

<table>
<thead>
<tr>
<th>Plant</th>
<th>Data</th>
<th>Model</th>
<th>Application</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eggplant</td>
<td>RGB images</td>
<td>CNN</td>
<td>Pollen development stages identification</td>
<td>[35]</td>
</tr>
<tr>
<td>35 rare species</td>
<td>Composite images (12 channels)</td>
<td>CNN</td>
<td>pollen classification</td>
<td>[36]</td>
</tr>
<tr>
<td>Maize</td>
<td>RGB images</td>
<td>CNN</td>
<td>Stomata detection and classification</td>
<td>[37]</td>
</tr>
<tr>
<td>Lenga beech</td>
<td>RGB images</td>
<td>CNN</td>
<td>Xylem vessels segmentation</td>
<td>[38]</td>
</tr>
<tr>
<td>Plant cell datasets</td>
<td>Confocal laser scanning microscope imagery</td>
<td>CNN</td>
<td>cell segmentation and tracking</td>
<td>[39]</td>
</tr>
<tr>
<td>Crambe</td>
<td>X-ray images</td>
<td>CNN</td>
<td>Seed quality assessment</td>
<td>[31]</td>
</tr>
<tr>
<td>Okra</td>
<td>Hyperspectral images</td>
<td>CNN</td>
<td>Hybrid seed classification</td>
<td>[40]</td>
</tr>
<tr>
<td>Chickpea</td>
<td>RGB images</td>
<td>CNN</td>
<td>Seed classification</td>
<td>[41]</td>
</tr>
<tr>
<td>Four plant datasets</td>
<td>RGB images</td>
<td>CNN</td>
<td>Species classification</td>
<td>[42]</td>
</tr>
<tr>
<td>Sorghum</td>
<td>RGB images</td>
<td>CNN and LSTM</td>
<td>Nitrogen deficiency stress</td>
<td>[43]</td>
</tr>
<tr>
<td>Potato</td>
<td>RGB images</td>
<td>CNN</td>
<td>Tuber disease detection</td>
<td>[44]</td>
</tr>
<tr>
<td>Potato</td>
<td>Hyperspectral images</td>
<td>1D-CNN</td>
<td>Anthracnose identification</td>
<td>[45]</td>
</tr>
<tr>
<td>Soybean</td>
<td>Hyperspectral images</td>
<td>3D CNN</td>
<td>Charcoal rot disease identification</td>
<td>[30]</td>
</tr>
</tbody>
</table>

3.3. Population level

Population plant science refers to the study of plant communities, whether in natural habitats (e.g., forests, grasslands, and deserts) or in agricultural land. Population studies tackle various issues such as plant cover classification and surveys, plant communities’ biological functions, the interactions between individuals or species, and how plant populations are being influenced and influence the surrounding environment (Table 3).

Spatial imagery is the predominant study material in DL population studies. The exponential increase in the availability of remote sensing (RS) datasets increased DL dependant plant population studies. Furthermore, the recent developments in unmanned aerial vehicles (UAV) manufacturing resulted in the production of lighter, more powerful, and affordable drones capable of carrying all sorts of image acquiring systems. These developments created a precious opportunity for research teams to create their RS datasets.
Table 3. Recent deep learning applications in plant population studies.

<table>
<thead>
<tr>
<th>Population type</th>
<th>Data</th>
<th>Model</th>
<th>Application</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wetland</td>
<td>Hyperspectral images</td>
<td>CNN</td>
<td>Plant Communities mapping</td>
<td>[46]</td>
</tr>
<tr>
<td>Forest</td>
<td>RGB images</td>
<td>CNN</td>
<td>Tree classification</td>
<td>[47]</td>
</tr>
<tr>
<td>Cornfield</td>
<td>RGB images</td>
<td>CNN</td>
<td>Corn plant count and plantation rows detection</td>
<td>[48]</td>
</tr>
<tr>
<td>Citrus fields</td>
<td>Hyperspectral images</td>
<td>CNN</td>
<td>Citrus phenotyping</td>
<td>[49]</td>
</tr>
<tr>
<td>New Zealand glacier foreland</td>
<td>RGB images</td>
<td>CNN</td>
<td>Mapping plant cover</td>
<td></td>
</tr>
<tr>
<td>Central Chile</td>
<td>RGB images</td>
<td>CNN</td>
<td>Mapping invasive species</td>
<td>[50]</td>
</tr>
<tr>
<td>New Zealand primary forests</td>
<td>RGB images</td>
<td>CNN</td>
<td>Mapping vegetation succession</td>
<td></td>
</tr>
<tr>
<td>Central Chile</td>
<td>Hyperspectral images</td>
<td>CNN</td>
<td>Segmentation of plant species</td>
<td>[51]</td>
</tr>
<tr>
<td>Soybean field trials</td>
<td>Hyperspectral images</td>
<td>CNN</td>
<td>Predicting days to maturity, plant height, and grain yield</td>
<td>[52]</td>
</tr>
<tr>
<td>Soybean breeding line</td>
<td>1D and 2D fusion input</td>
<td>Faster R-CNN and YOLOv4</td>
<td>Yield prediction under drought stress</td>
<td>[53]</td>
</tr>
<tr>
<td>Forest</td>
<td>Hyperspectral images</td>
<td>Faster R-CNN and YOLOv4</td>
<td>Pine wilt disease early detection</td>
<td>[54]</td>
</tr>
</tbody>
</table>

4. Current and future aspects of deep learning in plant science

Although DL can provide viable solutions to many complicated issues in plant research, the application of DL in plant science studies is still faced with many obstacles. One of the hurdles facing DL in plant science is dataset size and availability. Due to the relatively high costs of omics research, and the labors related to all three discussed levels of plant research, constructing largely enough datasets is a tiresome task. In fact, DL models depend greatly on the size and balance of the training and validating datasets to illustrate adequate generalization since small datasets usually result in overfitted models [55]. To overcome this obstacle, image augmentation techniques (Fig. 3) are used to expand the dataset size. Various researchers employed data augmentation to increase the accuracy of the developed DL models to some extent. However, increasing dataset size by adding new samples is still required to obtain more reliable results, especially in omics studies, where augmentation techniques are inapplicable.

![Figure 3. Image augmentation strategies. (A) is the original image, while (B-E) represent various augmentation techniques applied to the same image, such as mirroring (B), rotating (C), stretching (D), and transitioning (E).](image-url)

On the other hand, large datasets require long hours of human supervised labeling since the training process requires adequately labeled data. However, the current rise of self-supervised learning (SSL) and semi-supervised learning might provide a viable solution to this problem since these techniques require unlabeled or partially labeled datasets to learn [56].

Among all three discussed levels of DL research in plant science, omics are still poorly represented. This poor representation is mainly due to the high costs of datasets generating and the unsuitability of omics data in its raw forms to be used in DL training which requires long hours of processing. Therefore, developing new methods for omics data preparation and pre-processing is necessary. Furthermore, employing layer visualizing methods such as...
saliency map [23] and feature map [27] might be of great importance. These maps provide valuable information regarding the features CNNs use to classify and predict. Therefore, these visualizations can assist omics research in selecting the best plant characteristics.

As for micro/macrosopic levels, it is expected that DL techniques will transform from a complementary assistant tool to play more vital roles. Novel DL models are being developed daily to provide a deeper understanding of physiological processes and biological interactions between plants and their abiotic and biotic surroundings [43-45]. Additionally, DL has excellent potential to provide rapid and accurate judgments in agricultural production lines [27].

Population level DL studies are expected to play significant roles in the real-time tracking of invasive species and plant population dynamics in natural habitats [46][47][50][51]. Furthermore, the new models will introduce a new age of cost-effective and accurate DL-assisted agricultural extension [52][53], which will significantly positively impact food production chains in the near future.

References


43. Liu F, Xiao Z. Disease Spots Identification of Potato Leaves in Hyperspectral Based on Locally Adaptive 1D-CNN. In2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA) 2020:355-358. DOI


