Urban Food Systems Symposium

Overcoming AI Model Training Barriers Towards Autonomous Controlled Environment Agriculture Systems

Shing I. Chang
Kansas State University, changs@ksu.edu

Michael Prichard
Kansas State University

Xiaolong Guo
Kansas State University

Lizhi Wang
Oklahoma State University

See next page for additional authors

Recommended Citation

Chang, Shing I.; Prichard, Michael; Guo, Xiaolong; and Wang, Lizhi (2024). "Overcoming AI Model Training Barriers Towards Autonomous Controlled Environment Agriculture Systems," *Urban Food Systems Symposium*. https://newprairiepress.org/ufss/2024/proceedings/1

Abstract

Vertical farming (VF) is a type of Controlled Environment Agriculture (CEA) that promises high yield, year-round production, reduced land usage, shorter supply chains, and sustainable agriculture practices. However, its operation is a challenge to small farmers. Recent artificial intelligence (AI) advances have provided opportunities to design autonomous CEA systems. Unlike classic automation, in which each component is set to operate according to a timer, the proposed autonomous system coordinates all system components based on a set of goals (e.g., yield, taste, and energy consumption). The proposed autonomous system aims to model, monitor, and control the CEA ecosystem based on big data gathered in a proposed agriculture digital twin called Agri Generative Digital Twin (AGDT) platform in the cloud. Digital twin concepts have been applied in various manufacturing applications where data collected via sensors (i.e., in the context of IoT Internet of Things) resides in the cloud. Manufacturers can ask what-if questions to forecast system performance and simulate potential production changes without interrupting the physical production systems. Similarly, the proposed AGDT would collect growing data from crops and the environmental factors and allow farmers to ask VF operational questions. Collecting data from various VF systems and incorporating transfer learning techniques in AI will meet the substantial data demands for training AI models. The trained AI models can design optimal growing CEA recipes based on multiple objectives such as crop yields, quality, and costs. Farmers/users of the proposed autonomous AI models could optimize growth production and reduce costs to generate steady incomes, thus achieving sustainable agriculture practices.

Keywords

Vertical farming (VF), Controlled Environment Agriculture (CEA), Artificial Intelligence (AI), Machine Learning (ML), Digital Twin (DT), Autonomous System

Disciplines

Computer Sciences | Data Science

Presenter Information

Shing I. Chang, Michael Prichard, Xiaolong Guo, and Lizhi Wang

Overcoming AI model training barriers towards autonomous controlled environment agriculture systems

ABSTRACT

Vertical farming (VF) is a type of Controlled Environment Agriculture (CEA) that promises high yield, year-round production, reduced land usage, shorter supply chains, and sustainable agriculture practices. However, its operation is a challenge to small farmers. Recent artificial intelligence (AI) advances have provided opportunities to design autonomous CEA systems. Unlike classic automation, in which each component is set to operate according to a timer, the proposed autonomous system coordinates all system components based on a set of goals (e.g., yield, taste, and energy consumption). The proposed autonomous system aims to model, monitor, and control the CEA ecosystem based on big data gathered in a proposed agriculture digital twin called Agri Generative Digital Twin (AGDT) platform in the cloud to mitigate data challenges in AI modeling training. Digital twin concepts have been applied in various manufacturing applications where data collected via sensors (i.e., in the context of IoT Internet of Things) resides in the cloud. Manufacturers can ask what-if questions to forecast system performance and simulate potential production changes without interrupting the physical production systems. Similarly, the proposed AGDT would collect growing data from crops and the environmental factors and allow farmers to ask VF operational questions. Collecting data from various VF systems and incorporating transfer learning techniques in AI will meet the substantial data demands for training AI models. The trained AI models can design optimal growing CEA recipes based on multiple objectives such as crop yields, quality, and costs. Farmers/users of the proposed autonomous AI models could optimize growth production and reduce costs to generate steady incomes, thus achieving sustainable agriculture practices.

Key Words: Vertical farming (VF), Controlled Environment Agriculture (CEA), Artificial Intelligence (AI), Machine Learning (ML), Digital Twin (DT), Autonomous System

INTRODUCTION

The United Nations' World Food Programme (WFP) currently estimates that nearly one billion people across the globe are food insecure (FAO et al., 2020). Data continues to illustrate a growing food crisis in both underdeveloped and developed parts of the world (World Bank, 2021; Hincks, 2018). Factors contributing to this crisis include severe weather, supply chain difficulties, wars, geopolitical struggles, COVID-19-related disruptions, etc. A viable solution to this challenge is the implementation of vertical farming systems integrated near large urban populations, large food processing facilities, or climate sensitive areas not conducive to traditional methods of agriculture. This method of farming mainly grows food indoors vertically using novel growth media (on vertical surfaces or multiple platforms of various heights) while also using a mineral nutrient water solution rather than growing crops on traditional farmlands. Vertical farming implemented close to major cities promises short food miles that minimize supply chain disruptions.

VF methods are not without their own challenges. Besides high setup costs, one of the main issues is the high energy consumption that leads to high operation costs. As opposed to traditional farming where the sun provides free photosynthesis light to crops growing outdoors, vertical farming requires mechanically controlled indoor climate for lighting, temperature, humidity, water recycling, and CO_2 supply. Albright, a CEA pioneer, warned about the high cost, large energy use, and giant carbon footprint (Shackford, 2014). This manuscript aims to answer the following research questions: "How do we co-design an autonomous artificial intelligence (AI)/ machine learning (ML) system for vertical farming based on plant monitoring and environmental sensing? How can we overcome the barriers for training the proposed AI/ML models? ”

Many large-scale VF operations are successful. Examples include facilities in Massachusetts, UK, and Japan. These operations carry high set-up costs due to automation set-up costs and large-scale indoor facilities The operational costs associated with large-scale operations are also prohibitive for small farmers to enter the otherwise profitable VF operations. One of the main cost items is the electricity used for artificial lights. An affordable, small-scale VF system is the use of shipping containers. A variant of this idea combines shipping containers and greenhouses together. For example, a greenhouse can be built on top of a 12.1 x 2.43 x 2.59 (meter, L x W x H) container. All CEA mechanics are housed in the container while the plants grow in the green house. This kind of setup leverages sunshine to mitigate the use of electricity for lighting. Automation is often implemented in such a system, in which each component is set to operate according to a timer. An example of an automation system is a traditional sprinkler system for a lawn. It usually turns on all stations sequentially each for half an hour starting at 4 AM Monday, Wednesday, and Fridays. However, this type of automation may not optimize results in terms of costs and yields. Regardless of large or small CEA operations, the lower the operational costs, the better the profits and potential benefits to the environment. In this sprinkler system analogy, a schedule water cycle should be suspended if there is high chance of rain for the day. If the rain does not come, an autonomous system should be able to adjust subsequent water schedules to make up for the missing cycle.

Traditional empirical modeling approaches, such as the response surface methodology (RSM) (Myers and Montgomery, 2016), may be a candidate for a descriptive model that establishes the relationship between CEA parameters and responses, such as yield. Usually, RSM models are polynomial of CEA variables, such as lighting, water frequency, nutrient amount. RSM models may not accommodate more complex input, such as aerial and root images, and greenhouse temperature cycles during a growing cycle. In addition, RSM models are usually applied when the final outcomes such as the yield are realized. In this regard, AI technologies may provide a solution for advancing predictive models into prescriptive models.

Recent advances in AI and ML have provided opportunities to design an autonomous VF system where all system components are coordinated, based on a set of goals (e.g., yield, taste, and energy consumption). In the sprinkler example, an autonomous system would only turn on the sprinkler stations based on inputs from soil moisture sensors and weather forecast. The proposed autonomous system aims to model, monitor, and control the vertical farming ecosystem based on the big data gathered. Specifically, the proposed algorithms would integrate plant science, such as photosynthesis limiting factors and a data-driven AI model for an optimal growing CEA recipe based on multiple objectives such as crop yields, quality, and costs. Information gleaned from the crop-growing images provides the base of climate control, lighting control, fertilizing, and pest control. Farmers that can operationalize an “artificially intelligent big data ecosystem” would be able to optimize growth production and reduce costs to generate steady incomes, and thus achieve sustainable agriculture practices. For example, an autonomous VF system will minimize electricity consumption through data from light sensors and images from leaves and roots during a growing cycle.

One of the main barriers of adapting AI for any problem is training, which requires a large amount of data. For small-scale VF operations, it is difficult if not impossible to collect enough data for AI model training. This manuscript will tackle this issue using a concept called digital twins (DT) originated from manufacturing (He and Bai, 2021; Fumagalli and Macchi, 2017). The core idea is to pool data from multiple VF operations to achieve critical amount of data for AI model training. Then each VF “pod” can download the trained model to allow further customized training and usage. The rapid advances in big AI platforms from leading chip makers such as Nvidia (Wong, 2024) further support the vision of the proposed Agri-Generative Digital Twin (AGDT) framework.

MATERIALS AND METHODS

The Proposed Digital Twins for Autonomous VF Operations

Data collected in the hydroponic VF containers and from plant sensors and cameras will establish a data stream for AI/ML model training and plant physiology studies. For model training purposes, a stream of crop growth (via images), crop nutritional values, energy consumption, and CEA parameter data need to be collected. The proposed framework collects the data locally first and then uploads it to the data center in the cloud.

The proposed AGDT hosts a mathematical representation of the complex interactions between a crop and its growing environment at the physiology level, which provides domain knowledge to support agricultural stakeholders. Details of the proposed AI models for this mathematical representation will be discussed in the following sections. The proposed AGDT as shown in Figure 1 has the physical parts with their digital counterparts. The digital collection of all physical parts consists of several main components: (1) data storage (database & file system): CEA parameter data and plant images, (2) predictive and prescriptive AI models, (3) Large Language Models (LLM) hosting plant science knowledge, (4) user feedback data, and (5) user interface. The concept of digital twins derives from manufacturing operations where data generated from sensors scattered throughout a factory is collected in a cloud platform. A virtual or digital copy via sensor readings replicates physical factory operations. In other words, this digital replicate is the digital twin of the physical factory. A digital twin allows simulated changes in the virtual factory without interrupting production. The proposed AGDT adopts this concept with the application for CEA operations. Thousands of VF operations may reside in AGDT to share data for AI models for common use. The individual data stream from each VF system through the trained data provides customized solutions. An analog is a self-driving AI model trained on driving data from millions of cars, but each self-driving car navigates to different roads and destinations using the same model. The input data of a self-driving AI model is images from multiple cameras while that of a CEA AI model includes CEA parameter readings and images from crops during a growing cycle. VF pods or farms in the proposed AGDT network may contribute to large numbers of growing data that overcomes the data challenges in AI model training.

Various data driven or process-based models have been proposed for crop growth analysis and prediction (Khaki and Wang, 2019; Khaki et al., 2020). The proposed new modeling and solution techniques, on the other hand, focus on adjusting CEA parameters during a planting cycle so that the desired outcomes can be achieved. The following sections describe both predictive and prescriptive modeling approaches for VF operations.

Predictive And Prescriptive Modeling Approaches

Figures 2a and 2b show two general system modeling frameworks. Figure 2a shows a predictive model framework that establishes relations between process variables and response variable(s). A simple example is a regression model $\boldsymbol{y}$=f($\boldsymbol{x}$) where $\boldsymbol{x=(x_1, x_2, \ldots)}$ is a vector of input (or independent) variables, and $\boldsymbol{y}$ is a vector of output (or dependent) variable(s). Note that multiple output variables are possible by multiple regression polynomials for each y variable. One of the output variables may be the crop yield, and another output may be the crop quality. This approach has a major drawback: some output variables, such as yield, are often evaluated at the end of the growing cycle. The input variables are usually a constant setting. For example, $\boldsymbol{x_1}$ is the variable for light intensity in three levels: low, medium, or high. This simple structure may not reflect the reality in a CEA operation because the light intensity may change during a growing cycle.

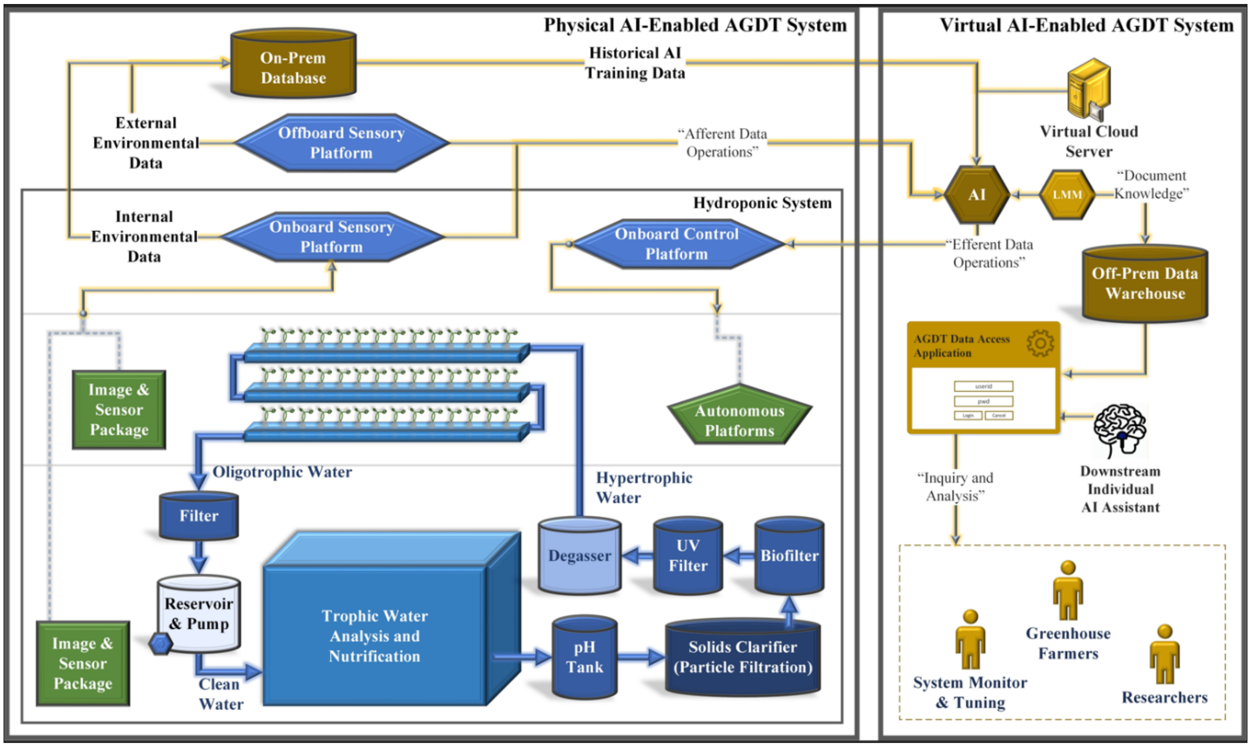

Figure 1. The Architect of the Proposed AGDT

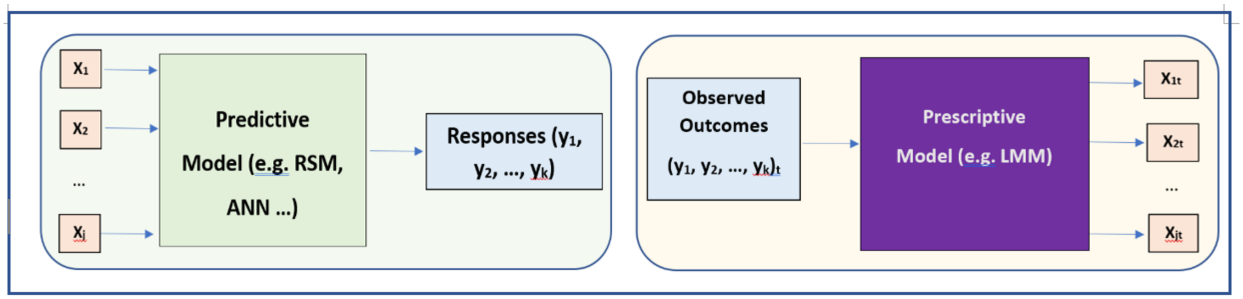
X_1
X_2
...
X_j
Predictive Model (e.g. RSM, ANN ...)
Responses (y_1, y_2, ..., y_k)
Observed Outcomes (y_1, y_2, ..., y_k)$_t$
Prescriptive Model (e.g. LMM)
X_{1t}
X_{2t}
...
X_{jt}

Figure 2. (a) Predictive Model. (b) Prescriptive model

Figure 2b shows the opposite approach to seek settings for ***x*=(*x*$_1$, *x*$_2$, ...)** given the objective ***y*,*** which is the desired outcome. An inverse transformation may be used to obtain the estimated input setting in simple cases such as regression. However, this simple approach is not applicable to a CEA system because the function is time dependent. Specifically, the CEA variables and the functions leading to the desired outcome ***y**** are time dependent. It is a challenge to model the relationship between CEA operations and the outcomes properly since the CEA setting changes over time. The following section describes a machine learning approach to achieve stage-wise control, assuming CEA setting within each stage is homogeneous (i.e., stay the same). We also assume that the proposed digital twin for VF operations will collect enough data and the planning cycle for a crop can be segmented into homogeneous stages.

A Predictive Modeling And AI Framework For Crop Yield Estimation

For predictive modeling, deep neural networks (DNN) such as convolutional neural networks (CNN) and recurrent neural networks (RNN) are capable of predicting crop yield. For example, Khaki and Wang (2019) reported the use of DNNs for Syngenta Crop Challenge of 2,267 maize hybrids planted in 2,247 locations between 2008 and 2016. The proposed DNN models outperformed traditional models such as Lasso, shallow neural networks, and regression tree. Khaki et al. (2020) have further considered environmental data and management practices in a CNN-RNN model for crop yield predictions. The proposed model captures time dependencies of environmental factors and genetic improvement of seeds over time due to the use of RNN. The proposed framework can be summarized in Figure 3, where CNN-W is used to account for weather, and CNN-S is used for soil impact. The data for training contained average corn and soybean yield between 1980 and 2018 across 13 states. The symbol M represents management data, including the weekly cumulative percentage of planted fields within each state, starting from April every year. Data from 2008 to 2015 was used to predict yields for the years 2016 to 2018. While the proposed CNN-RNN model substantially outperformed other models, such as DNN and Lasso, the yield predictions were meant for annual forecasts. Since the CEA parameters can be manipulated, we will explore the possibility of adjusting these parameters in the next section for better yields and good crop quality during a growing season/cycle.

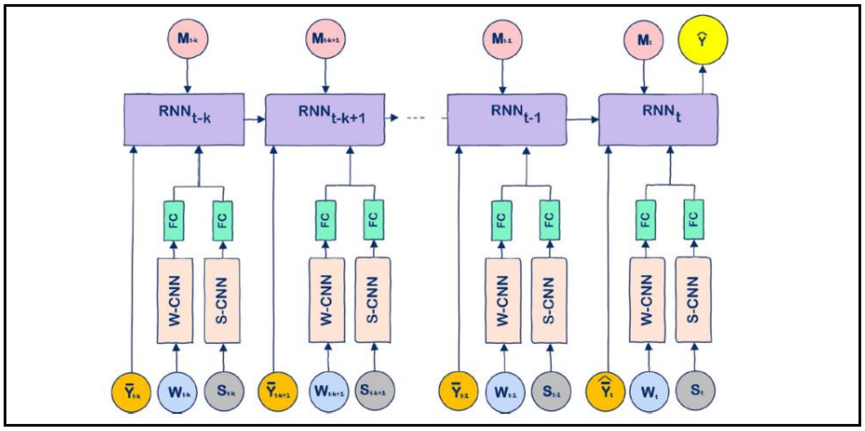

Figure 3. A CNN-RNN Framework to Crop Yield Prediction (Ref. Figure 2 of Khaki *et al.*, 2020)

A Predictive Modeling And ML Framework For Mid-Growing-Cycle Adjustments

One of the crucial features of the proposed autonomous VF system is the ability to adjust operational parameters, such as the frequency and duration of lighting, so that the desired yield and crop quality, such as taste, can be achieved. In this case, the output variables are expanded to more than just numerical types, such as yield. We propose to include aerial and root images as surrogate estimates for where t=1,2, …, n for n stages in a grow cycle. Based on the images of crop growth, a forecasting model (i.e., a descriptive model) can be established for mid-growing-cycle yield prediction. Recent studies have also concluded that specialized LED with different spectra and variable grow cycles can improve crop taste (Zhang *et al* 2019). In addition, other environmental factors, such as room temperature and water nutrients – all contributing to the yield and crop nutritional quality.

Our previous work on a laser power bed fusion (LPBF) manufacturing process has demonstrated using an AI/ML framework to monitor 3D printing process parameters and print quality (Amini and Chang, 2018). The proposed AI/ML modeling framework is called Multi-Layer Classification for Process Monitoring (MLCPM), as shown in Figure 4, in which clustering and feature selection methods were integrated to achieve dimension reduction and satisfactory prediction. Assessment of print quality is based on layer-by-layer, top-view images. Process data from multiple 3D printing machines producing the same part are pooled to expedite data collected for model training. This same framework can be applied to the proposed VF application in which daily image data from different VF containers growing the same crop can be pooled and clustered to identify crucial moments (i.e., stages) for crop prediction, as shown in multiple cycles in Figure 4. Note that M is a very large number.

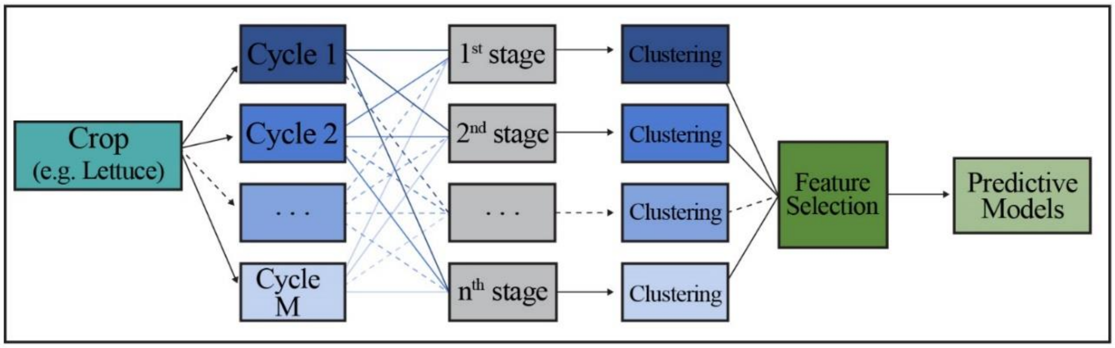

Figure 4. A Multi-Stage Classification Process Prediction (MLCPM or MSCPP)

The proposed AI/ML modeling framework adopted for VF applications is called Multi-Stage Classification for Process Prediction (MSCPP). For the proposed CEV/VF application, data X_{ijk} from various sensors represents the collected data for growing cycle ***i***, stage ***j,*** and sensor ***k***. The outputs of each growing cycle are denoted as $\mathbf{y_{io}}$ where ***o***=yield, nutrition, taste, energy consumption, etc. We propose to use a clustering algorithm $g(.)$ to group all growing cycles into a few growing recipes that contain all environmental factor readings in all growing stages. Since the number of the factor combinations is huge, this clustering operation would achieve dramatic dimension reduction. The outcome is $C_j = g(X_{ijk}), C_j \in \left\{C_1, C_2, \dots, C_{p_j}\right\}$, j=1,2,...,n, where the number of clusters is $\boldsymbol{p_j}$ for the j^{th} stage. For notational simplicity, the cluster outcome is meant for one growing output (e.g., ***o***=yield). Then, a second clustering operation f(.) is applied to the outcome of the first cluster operation so that the predicted model for each output ***o*** can be established at each significant stage. In general, assuming n significant stages, the j^{th} classification model can be built as a set of equations: $\widehat{Y_o} = f_j\left(C_1, C_2, \dots, C_j\right),\ j = 1,2, \dots, n$. Once the function fj() is established, we would like to seek for an appropriate adjustment action pending on the prediction outcome. To achieve this task, we use an ensemble of prescriptive models and ML algorithms to seek x_t^* at time t where t is at a crucial stage.

The process that leads to determining the most significant stages and process parameters affecting the outcome, such as yield, is called **feature selection**, which enables the proposed autonomous system to adjust the mid-growing cycle adjustment. For example, model l=2 represents the second significant stage in the growing cycle of iceberg lettuce. All the data collected from multiple growing cycles and containers determines this stage. At this stage, we can compute the projected yield and open the door to search for possible future process parameters for mid-cycle adjustment. Since all process parameter combinations have been clustered and associated with actual yields, we can use a search algorithm to find growing cycles with desired yields. Then, the process parameter combinations from the second significant stage to the third significant stage associated with cycles with the desired yields can be gleaned for process control. The feature selection step can be skipped if the number of stages is small. In this case, an adjustment decision should be made at the end of each stage. The proposed algorithm is data-driven. Therefore, the more data with successful crop yields, the better the performance of this autonomous framework.

One of the main challenges in the proposed MSCPP framework is the estimation of intermediate outcomes. In the case of crop yield, how can it be forecasted before a harvest is accomplished? Suppose we collect daily images of leaves and roots from a sample crop such as lettuce. We would like to predict the crop yield or food quality at each critical growing stage. Figure 5 shows the proposed framework. Specifically, all images leading to the end of a growing stage can form a time-lapse video I_t. For example, the rosette is a critical stage from the plant developing its first true leaves to the leaves clustered tightly together near the soil and at a similar height. We need to collect a large number (M) of time-lapse videos for model training. In this case, a deep learning model can be trained to estimate crop yield or quality based on the daily time-lapse videos up to the rosette stage. Since the training data set is based on crop cycles that have already been harvested, we can label the outcome y_t as either good or bad as a binary classification problem. Alternatively, a score between 0 and 100 can be used. Since there are multiple stages, we need to train multiple classification models – each for a stage: $\widehat{Y_o} = h_j\left(I_1, I_2, \dots, I_j\right),\ j = 1,2, \dots, n.$ where o is the index for outcome such as yield or quality and j is the index for stage. The function $h_j(.)$ is the video classifier for stage j. Images I_j from all stages are non-overlapping time-lapse images (i.e., video) of growing crops.

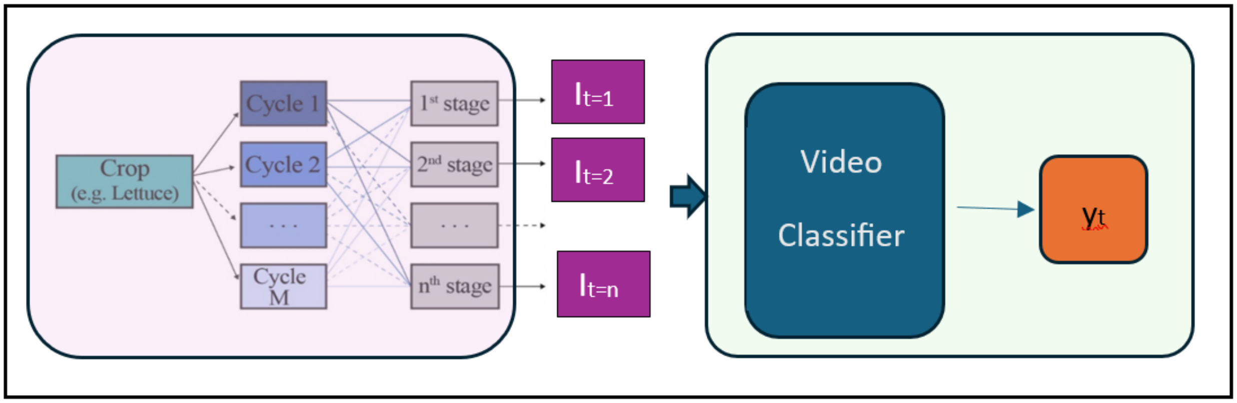

Figure 5. A Proposed Video Classifier for Forecasting Crop Outcomes

A Prescriptive Generative AI Framework for Mid-Growing-Cycle Adjustments

Since the publication of the Google landmark paper titled "Attention is all you need" in 2017 (Vaswani *et al.*, 2017), the self-attention mechanism in a transformer has had great impacts in many fields, such as Open AI's Chat-GPT and Tesla's Full Self Driving (FSD version 12 and after). Chat-GPT 3.5 and 4 are based on a Large Language Model (LLM) mainly for text-related applications, while the inputs for FSD are videos from multiple cameras. Dosovitskiy et al., (2020) extended the self-attention mechanism to image classification, such as "An Image is worth 16x16 words", where an image is divided into 16x16 stamps, then strung together like a sentence of 16x16 words. The recent introduction of Chat-GPT 4o (GPT-4o, 2024) has further incorporated images, audio, and videos in addition to text as both inputs and outputs. Google has also announced its Large Multi-modal Model (LMM) in its AI Gemini release (Pichai and Hassabis, 2024). Chen and Ho (2021) demonstrated a model called MM-ViT, combining both compressed video and audio inputs for action classification.

The proposed prescriptive modeling framework is similar to FSD in that the outputs of FSD are turning the wheel left or right, increasing or decreasing acceleration, and applying break, while the CEA outputs include increasing or decreasing lighting, increasing or decreasing nutrition, increasing or decreasing water flow/frequency, … etc. Specifically, the prescriptive model will also need information on the current CEA parameter setting, application history, and a stream of aerial and root images. The following section would detail how the adjustment works.

The proposed prescriptive DNN framework shown in Figure 6 is capable of mid-growing-cycle adjustments in one integrated model. A generative AI model via the self-attention mechanism is adopted to model the relationship between inputs taken over time and CEA adjustments. Specifically, the inputs include aerial and root images I_t and current setting of the CEA variable vector $\boldsymbol{x}$***=(***$\boldsymbol{x_1}$***,*** $\boldsymbol{x_2}$***, …)*** and the outputs are the changes of CEA variable vector, Δ***x*** at time t***.*** We will use the same stage assumption in the predictive model framework, that is to break a growing cycle into *n* stages. Time-lapse aerial and root images I_t within a stage are then strung into a video. CEA parameters associated with each stage are collected in a vector x_t.

Through the self-attention mechanism, a score is generated for each element that feeds into a sequence of transformation operations that determines which CEA parameters would contribute the most to improving the outcome. Specifically, any input vectors feeding into a self-attention mechanism generate three components: Q, K, V via dot product operations, respectively. Then, self-attention scores are computed via another dot product operation. After a SoftMax operation normalizes each column of all dot product vectors, they are collected column-wise in the output matrix O. These output vectors in O from the self-attention box (e.g., the purple box in Figure 6) feed into a fully connected network (FCN) to generate Δ***x*** at time t***.*** Note that the weight parameters self-attention box (called a transformer) and FCN must be trained. All other matrices are provided. In general, a generative model requires more data to train than a CNN, a RNN (recurrent neural network), or a hybrid of CNN/RNN model. Since CNN does not carry memory, RNN architect is a way to string a series of CNN models where outputs of a prior CNN feed into its adjacent CNN to bring the memory required to solve a problem. The proposed generative model is more flexible in connecting all input vectors. In this case, the input vectors carry both the crop images and CEA parameters setting over time. This flexibility allows the early results to possibly affect later outcomes without making prior assumptions. The model performance depends solely on the training data rather than model assumptions!

The prescriptive modeling approach is much more straightforward than the predictive approach. Users/farmers would directly obtain an action plan in Δ***x*** at time t from the prescriptive model during the growing cycle. However, the power of the proposed prescriptive model comes at a cost because the model accuracy is directly correlated to the number of data needed for training. The data required for training generative AI models is far more demanding than that for predictive and ML models. This computational restriction will ease over time as the recent AI infrastructure investments from various tech giants such as Microsoft, Google, and Meta (Wong, 2024) will provide amble computational resources. The computational woe of AI model training is no longer a constraint. We predict that the bottleneck of the proposed prescriptive CEA models is not the computational hardware but the data available for training.

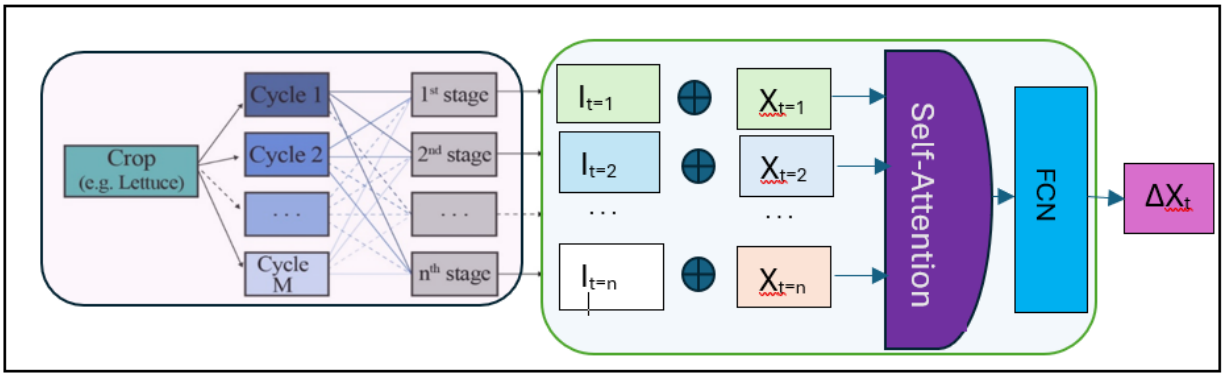

Figure 6. The Proposed Self-Attention Mechanism in DNN for Mid-Cycle Adjustments

A Generative AI Learning Framework Integrating Existing Plant Science Knowledge, CEA Data, And Farmer Feedback

Since the pure data-driven approaches described in the previous sections require a huge amount of data, how do we transition into autonomous VF operations before the prescriptive models are accurate enough? One plausible approach involves transferring existing knowledge into an AI model. Specifically, we propose to extract existing growing recipes, farming practices, and common plant science knowledge. Such knowledge may come from existing Scopus in textbooks and literature, e.g., the efficiency of photosynthesis under different temperatures, light intensity, and carbon dioxide concentration conditions, evapotranspiration rate under various temperature, humidity, and vapor pressure deficit environments. The Farquhar photosynthesis model (Farquhar et al., 2001) and the Penman-Monteith evapotranspiration equation (McColl, 2020), among many other models, belong to the category of process-based models. They are insightful but do not have good prediction accuracy. This knowledge provides general guidelines but does not apply to a specific CEA environment for mid-growing adjustments. Another approach is the use of data-driven components for photosynthesis and evapotranspiration, which are less descriptive but more predictive. Although, at the crop level, the hybrid model is still reasonably explainable, the key is to strike a good balance between explainability and prediction power. The learning outcome will be the transferable knowledge that can be used to accurately predict the growth of other crops and varieties with limited data sets.

LLMs or LMMs (Large Multimodal Models) can extensively leverage the knowledge conveyed and utilized through natural language and data for agricultural processes, in this case, the VF operations. We propose a multimodal AI assistant that integrates a natural language model akin to ChatGPT. This AI assistant will serve as a communication bridge between predictive AI models and farmers, who are the end-users of the envisioned VF container system. As illustrated in Figure 7, the proposed multimodal approach integrates natural language, aerial and root image data, and CEA data into a unified framework. This innovative communication platform will facilitate the exchange of insights and recommendations, enhancing the efficiency and effectiveness of urban farming practices. The centerpiece of this model is an AI assistant underpinned by an LMM (Yu, et al., 2023; Shukor et al., 2023). This AI assistant is a skilled conductor, unifying disparate modal inputs and blending language, visual cues, and data from in-field sensors into a cohesive information flow. The proposed AI assistant can deduce knowledge beyond its initial training dataset by interacting with downstream individual AI models, such as crop yield and food quality prediction models, as shown in Figures 5 and 6.

The intelligence of LLMs or LMMs is regarded as a compression of their training datasets, making these datasets the most crucial component of these models (Huang, et al., 2024). Figure 7 illustrates the four training phases of an LLM or LMM-based assistant: pre-training, supervised fine-tuning (SFT), reward modeling, and reinforcement learning. It details the algorithms, datasets, and representative tasks associated with each phase. The pre-training dataset is high-volume but low-quality, while the subsequent stages require high-quality human or agent-generated data. Data for these training phases are categorized into three types based on their source: common plant science knowledge (from existing Scopus in textbooks and literature), generated data from the proposed AGDT, and recorded agricultural data from farmers. The common plant science data is the largest in scale, followed by AI-generated data, with real-world recorded data being the smallest.

Moreover, we plan to balance the model explainability and predictive power of the proposed LMM framework. We will utilize general knowledge datasets as references for the general LMM, leveraging digital twin models to generate accurate data, enabling the LMM to provide explanatory information. While the raw data might be obscure to end-users, the LMM-enhanced synthesized data can reduce the barrier to utilizing this information sequentially. Appropriate classification and modeling are crucial preliminary steps for utilizing LMM-enhanced information for figures and chart data. Therefore, the domain-specific dataset construction and model training process requires participants with advanced backgrounds in both machine learning and plant science.

Specifically, in the pre-training phase of the proposed LMM, the raw textual content is used with a language modeling algorithm to create the base model. We will utilize the knowledge collected from common plant science in this phase. The supervised fine-tuning phase uses ideal responses to specific prompts as a dataset, producing the SFT model. The AGDT-collected data will be used to fine-tune the proposed model. The Reward Modeling phase introduces preference training through farmers' selections. The final reinforcement learning phase uses prompts and preference scoring to produce the RL (Reinforcement Learning) model. While Reward Modeling and RLHF (Reinforcement Learning from Human Feedback) are not identical, they are often discussed together due to their shared idea of enhancing performance. We plan to use real-world CEA data in these two stages.

By integrating this AI capability with the knowledge and operations of the proposed framework, we can streamline interactions for domain users, such as farmers and model builders. These users, who may not possess in-depth knowledge of CEA hardware and software, can easily convey their planting requirements to CEA experts using language, pictures, or both. The trained LMM may also generate insights in an easily digestible format, offering a clear overview of the system's real-time status to end-users. In addition to language generation, the proposed LMM can provide data inquiries similar to executing SQL commands but with more complex syntax. Finally, the task execution feature in LMM can serve as an interface for interactions between the external CEA environment and AGDT.

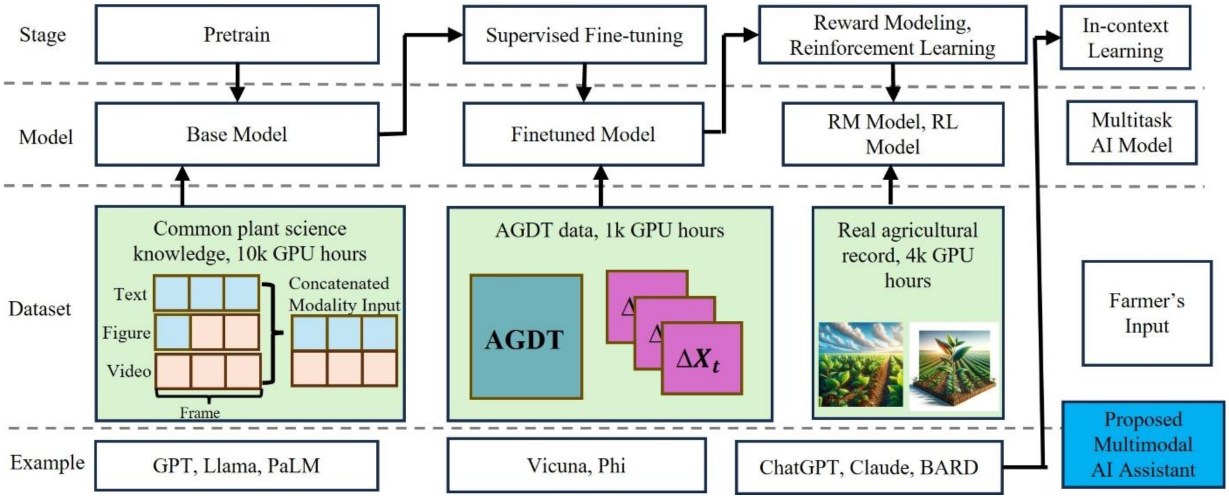

Figure 7. The Proposed Agriculture Domain-Specific Multimodal AI Assistant

RESULTS AND DISCUSSION

A successful autonomous VF system requires operational instructions during a growing cycle. The ultimate metrics include yields, quality, and operational costs. A desirable operation must achieve high yields and good-quality crops with minimal operational costs. How are these goals attainable? Assuming that CEA parameters and aerial/root images are collected constantly (e.g., hourly records for CEA parameters and daily pictures for images), the proposed models aim to provide CEA parameter adjustments throughout a growing cycle. It is an open-ended question of how often adjustments should take place. This is a design question that involves empirical evidence or experimentation. A plausible choice is a daily adjustment schedule that considers the images from the prior days. For example, different VF pods may be exposed to varying amounts of sunlight.

Systems in greenhouses may need less artificial lighting supplements on sunny days than on cloudy days. We will discuss the pros and cons of how predictive and descriptive modeling approaches may address this challenge in the following paragraphs with the focus on input data. Note that successful AI applications usually require thousands and millions of training sample data. In the following paragraphs, we will discuss how the proposed predictive and prescriptive AI approaches can successfully overcome the data barriers to achieve autonomous VF systems.

Data Requirements of Ensembled ML Methods vs. Generative AI Models

When a predictive approach such as the one shown in Figure 4 is used, each stage should accumulate enough changes for the clustering method to work correctly. In this case, the stages may be different growing stages, with many days within each stage. When the estimated model at a stage predicts a yield or quality shortfall, a search algorithm is activated for growing recipes that have similar CEA parameter settings up to the point of prediction and yet obtain desirable outcomes. Therefore, the adjustments depend on the recipes that were searched for and chosen. The main advantage of the predictive approach is that the data requirement for modeling and clustering is not as demanding as the generative AI model shown in Figure 6. Slight or moderate-size data (e.g., M= 30) is adequate. However, an empty search result may be possible. In such a case, users may not know the next step. Classical RSM predictive models won't accommodate images data. Typical clustering models such as K-Means do not require a large M either. In short, the ensemble approach of predictive models, clustering algorithms, and search algorithms does not require large amounts of trained data but its usage is limited and the outcomes may not be satisfactory.

Mitigating Data Requirement for CNN Models via Transfer Learning

The video classifier shown in Figure 5 requires a lot of data to train and perform properly. Usually, millions of images are needed for satisfactory accuracy. To address this challenge, a transfer learning approach may be used to alleviate the data burden. Specifically, a pre-train image model such as ResNet or VGG16 (Yan, 2023) for images from a public domain can be adopted as the base model. Typically, a 2D convolutional neural network (CNN) can be used to handle image inputs. For the video inputs, a 3D CNN architect (Tran, et al., 2017) can be used since the time dimension is a sequence of images over time. Therefore, a video input can be formulated as (2+1)D since 2D images plus the time dimension yield 3D. Then additional layers can be added to the base model. Training only takes place for the additional layers. Since the number of weights in these new layers is far less than the entire network, satisfying training results can be achieved with a much fewer number of images. Image data collected in the proposed AGDT will most likely be enough to train the proposed video classifier. Once the proposed video model is trained properly, feeding time-lapse images generated from the current growing cycles does not require modeling effort. Any autonomous VF users can feed their own aerial and root images and expect good prediction performance.

Digital Twin for Data Pooling Strategy

The data challenge for training the proposed prescriptive models shown in Figure 6 is immense. The significant advantages of the integrated prescriptive model come at the expense of prolonged training times, compounded by the need to collect large amounts of data. In this case, the number of growing cycles M is in the scale of hundreds of millions for good performance. The proposed AGDT is a good start for collecting all CEA parameters. When the number of users contributing to AGDT reaches a critical mass, it will reach the required large M number faster. An example is when Tesla released its FSD 12.3.x to millions of Tesla owners around March 2024 in North America. The driving data collected in a month exceeded the amount gathered from selected beta drivers over several years. Each subsequent FSD release has performed much better than their previous versions. On the same token, when the number of farmers subscribed to an AGDT reaches a critical mass, the training data will enable the successfully generative AI models envisioned in both Figures 6 and 7.

Integration of Human Language User Interface with Analytical AI Models

Before either the predictive or descriptive approaches are fully ready for truly autonomous VF operations, the proposed LMM generative AI models can navigate VF farmers with existing guidelines and plant science knowledge. Experienced farmers may not benefit as much as the novice farmers. The proposed LLM model within the LMM works like Chat-GPT. This LLM serves as a general help before the prescriptive models are fully trained. Farmers can ask any plant science or VF-related questions, and the trained model would summarize the questions asked. Farmers can then use this general information to manage their VF operations while generating CEA parameter data and crop images to be uploaded to the proposed AGDT in the cloud. After the performance of the prescriptive models reaches a satisfactory level, farmers can use the integrated LMM models for general knowledge and the prescriptive AI model in Figure 6 for daily operational adjustments.

CONCLUSION

We have proposed two AI/ML approaches for mid-cycle adjustments of VF operations. The predictive approach uses an ensemble of video classification via 3D CNN for crop outcome prediction, a cluster algorithm for dimension reduction by limiting the number of growing recipes, and a search algorithm for identification of successful paths to achieve desired yield and food quality. This approach requires much less data than the proposed descriptive approach. However, average VF system operators may not have the data science background to properly use the ensemble models and algorithms. On the other hand, the prescriptive models require massive data for training, with growing cycles numbering in the hundreds of millions for optimal performance. The output of this integrated model is the adjustment amount for each CEA parameter, which enables autonomous VF operations.

In the discussion section, we have outlined a couple of strategies in transfer learning and digital twins to mitigate the data barrios for AI model training. The proposed AGDT system will expedite data collection as user contributions increase, like Tesla's data collection strategy. Finally, generative AI models can provide interim support for VF operations by offering guidelines and plant science knowledge, particularly benefiting novice farmers. The proposed LLM can answer VF-related questions and assist in managing operations. After gathering sufficient data to train the proposed prescriptive AI models, the proposed LMM framework integrates the plant science LLM, predictive and prescriptive AI models, and feedback from farmers. This framework allows farmers to plan for VF operations, predict crop yield and food quality, and manage daily operational adjustments effectively.

For future research, costs of the operations are very important but not a part of the proposed prescriptive modeling framework described in this manuscript. For example, the rates of electricity may vary in a day or season. Optimal yields may not generate the most profits. The best growing recipes that are good for warm seasons may not be the optimal setting for the cold seasons. The proposed LMM model should be revised to accommodate all aspects of VF operations. Furthermore, since data collection is laborious and expensive, transfer learning from one crop to another may greatly reduce the model train effort. Specifically, if the DNN models from one crop can serve as the foundation (i.e. pre-trained) models for another crop. The number of samples needed will be greatly reduced. More research is needed to validate the performance of this approach.

ACKNOWLEDGEMENTS

The authors would like to thank the traveling support from the K-State GRIP funding to the 2024 Urban Food System Conference where this work was presented.

Literature cited

Amini, M. and S. I. Chang (2018). MLCPM: A process monitoring framework for 3D metal printing in industrial scale. Comp & Industrial Engineering, *124*, 322-330. https://doi.org/10.1016/j.cie.2018.07.041.

Chen, J., & Ho, C. M. (2021). MM-ViT: Multi-Modal Video Transformer for Compressed Video Action Recognition. *ArXiv*. /abs/2108.09322.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv*. /abs/2010.11929

FAO, IFAD, UNICEF, WFP, and WHO (2020). The state of food security and nutrition in the world report. DOI:10.4060/ca9692en, 2020.

Farquhar, G. D., S. Von Caemmerer, and J. A. Berry (2001). Models of photosynthesis. Plant Physiology, 125(1), pp. 42-45.

GPT-4o, (2024). Introducing GPT-4o and more tools to ChatGPT free users. https://openai.com/index/gpt-4o-and-more-tools-to-chatgpt-free/

He, B., Bai, KJ. (2021). Digital twin-based sustainable intelligent manufacturing: a review. Adv. Manuf. *9*, 1–21. DOI: 10.1007/s40436-020-00302-5

Hincks, J. (2018) The world is headed for a food security crisis. Here's how we can avert it. *Time*. Available: *https://time.com/5216532/global-food-security-richard-deverell/.*

Huang, Y., Zhang, J., Shan, Z., and He, J. (2024). Compression represents intelligence linearly. arXiv preprint arXiv:2404.09937.

Khaki, S. and Wang, L. (2019). Crop Yield Prediction Using Deep Neural Networks, Front. Plant Sci., *10.* https://doi.org/10.3389/fpls.2019.00621.

Khaki, S., Wang, L., and Arcontourlis, S. V. (2020), A CNN-RNN Framework for Crop Yield Prediction, Front. Plant Sci., 10. https://doi.org/10.3389/fpls.2019.01750.

McColl, K. A. (2020). Practical and theoretical benefits of an alternative to the Penman-Monteith evapotranspiration equation. Water Resources Research, 56(6). https://doi.org/10.1029/2020WR027106.

Myers, R and D. C. Montgomery (2016). Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 4th Edition, Wiley.

Pichai, S. and Hassabis, D. (2024) "Our next-generation model: Gemini 1.5," https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/.

Shackford., S. (2014). Indoor urban farms called wastful, 'pie in the sky', Cornell Chronicle. https://news.cornell.edu/stories/2014/02/indoor-urban-farms-called-wasteful-pie-sky

Shukor, M., Rame, A., Dancette, C., & Cord, M. (2023). Beyond task performance: Evaluating and reducing the flaws of large multimodal models with in-context learning. arXiv preprint arXiv:2310.00647.

Tran, D., Wang, H., Torresani, L., Ray, J, LeCun, Y, and Paluri, M. (2017). A Closer Look at Spatiotemporal Convolutions for Action Recognition, https://doi.org/10.48550/arXiv.1711.11248.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *ArXiv*. /abs/1706.03762.

Wong, Wylie (2024). Data Center Chips in 2024: Top Trends and Releases, Data Center Knowledge, https://www.datacenterknowledge.com/data-center-chips/data-center-chips-in-2024-top-trends-and-releases.

World Bank (2021). The world bank brief on food security and COVID-19. *https://www.worldbank.org/en/topic/agriculture/brief/food-security-and-covid-19*.

Yan, K. (2023). These are the 5 best pre-trained neural networks. https://medium.com/@kyan7472/these-are-the-5-best-pre-trained-neural-networks-23798e61a043.

Yu, W., Yang, Z., Li, L., Wang, J., Lin, K., Liu, Z., Wang, X. and Wang, L. (2023). Mm-vet: Evaluating large multimodal models for integrated capabilities. arXiv preprint arXiv:2308.02490.

Zhang, M., C. M. Whitman, and E. S. Runkle (2019). Manipulating growth, color, and taste attributes of fresh cut lettuce by greenhouse supplemental lighting, Scientia Horticulturae, *252*, 274-282. https://doi.org/10.1016/j.scienta.2019.03.051.