A review of Machine Learning applications in Controlled Environment Agriculture

Christos Charisis[1], Rudolf Erdei[2,3], Oliviu Matei[2,3], Pythagoras Karampiperis[4], and Dimitrios Argyropoulos[1]

[1] University College Dublin, School of Biosystems and Food Engineering, Dublin, Ireland
[2] Technical University of Cluj Napoca, North University Centre of Baia Mare, Romania oliviu.matei@cunbm.utcluj.ro
[3] HOLISUN, Romania rudolf.erdei@holisun.com
[4] SCiO, Greece pythagoras@scio.systems

1 Introduction

Controlled Environment Agriculture (CEA) has been used in various forms as a means of protecting crops from extremes of weather and other hazards. This protected environment also allows a space with precisely regulated environmental and cultural variables to produce crops in a more efficient way. The grower can maintain and control important crop-growth factors such as temperature, humidity, carbon dioxide, light, water and nutrients all year-round.

Apart from the widespread use of greenhouses [1], CEA is also known as indoor agriculture or vertical farming including soil-less farming techniques such as hydroponics [2], aeroponics [3] and aquaponics [4] with the possibility of increased yield by optimizing environmental conditions and inputs [5][6][7][8].

Considering the energy efficiency aspect, CEA still has not a clear advantage over conventional/open agriculture [6][7][9], but this can be partially attributed to the operation throughout the year in contrast with the periodical outdoor cultivation [9]. However, current research is focusing on developing new techniques and automation protocols utilizing the technological advance towards a more resource-efficient production [7][10], making CEA a more sustainable agricultural choice, for many crops e.g. tomato, lettuce, strawberry etc.

The concept of CEA derives from the usage of advanced technology e.g., sensors and actuators, which are located inside the structure to properly monitor the crop/microclimate interaction, control environmental parameters, and manage cultivation factors. Many different types of sensors and IoT implementations were reported in the recent literature [11][12][13][14][15]. In this way, data can be collected, analyzed and specific actions can be decided to change the conditions inside the controlled environment [10], leading to a better yield compared to respective outdoor cultivation [16].

Cultivation of mushrooms (e.g., *Agaricus bisporus, Pleurotus ostreatus, Lentinula edodes* among others) represents a special case of CEA, where agro-industrial residues are converted into high nutritious food under suitable environmental conditions of temperature, humidity and carbon dioxide. Several species of the genus Pleurotus are of particular interest since they can transform agricultural residues and agroforestry by-products into edible mushrooms. Nowadays, production of Pleurotus species amounts to ca. 30% of the total and corresponds to the fastest growing and most profitable section of the mushroom market during the last two decades [17]. Pleurotus spp. are commonly grown on pasteurized wheat or rice straw; however, they can be also cultivated on a wide variety of lignocellulosic substrates whose disposal is problematic [18]. The production involves four main steps: preparation of substrate, inoculation, incubation and fruiting, [18], that are highly dependent on environmental parameters, mainly temperature, humidity and CO_2 concentration in the growing rooms [19]. These factors exert a great effect on both growth and yield, thus their precise monitoring, prediction and control is of paramount importance along the mushroom production cycle from substrate preparation, to spawning and cultivation.

In recent years, technologies and tools used in agriculture are widely combined with, Internet of Things (IoT)[11], Big Data (utilizing Cloud Computing) [20] and Artificial Intelligence (AI)[21]. Emerging processing techniques adopt neural networks and deep learning approaches for providing new areas of application in the field of CEA [22]. The overall aim of all these developments is to provide new opportunities in managing crops and optimizing production processes. Technology can be applied for gathering data from vast networks in real-time, transmit these data to Cloud, develop large data banks, undertake data analytics to deliver mined and collated data to growers and other stakeholders in real-time.

A domain that excels in the presence of many available data is the domain of Machine Learning (ML). Many different models have been implemented so far, which are able to process data and analyze complex tasks in a better way than the commonly used simple sensors installations with human derived control rules or basic statistical models for the existing actuators. CEA installations can be regarded as very complex systems [23], where the use of more advanced mathematical algorithms in the form of ML models (e.g., Artificial Neural Networks, Random

Forest, Decision Trees, Support Vector Machines etc.), can be considered ideal for controlling the conditions of the environment, predicting the climatic variables behavior and increasing the produced yield.

The use of ML models is being explored in CEA for many years now. However, in recent years, the use of Deep Learning model in CEA applications is observed, as it can be seen in Figure 1. The aims of the researchers were to utilize these models to predict either the changes in climatic variables of the controlled environment or to monitor and predict the yield.

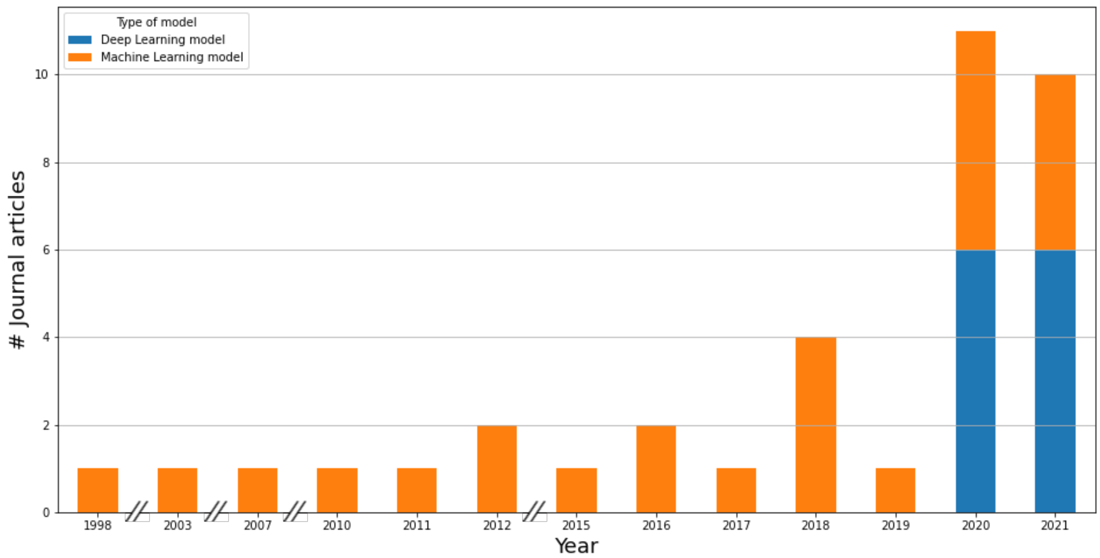

Fig. 1. Machine Learning vs Deep Learning studies per year in CEA bibliography.

Deep Learning models are more complex than the traditional ML models and require more resources and data to be implemented. The observed trend indicates that users have at their disposal larger datasets which give them the ability to train bigger and more complex models, in order to provide better solutions to the problems they try to solve. A workflow for a complete ML implementation on CEA is shown in Figure 2.

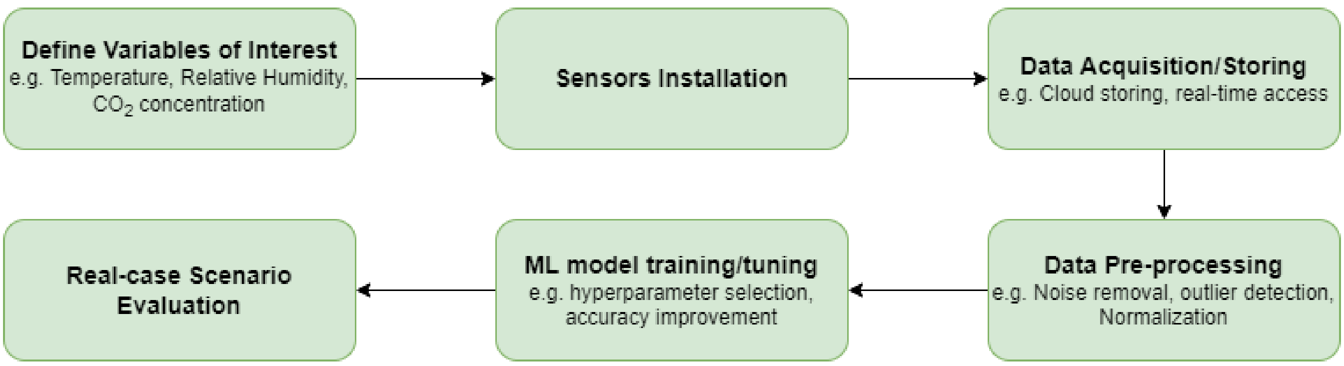

Fig. 2. Workflow for ML implementation in CEA.

Despite the increase in the use of ML models and the importance of mushrooms, to our best knowledge, a complete data-driven ML approach on mushroom cultivation does not. A very interesting and relevant attempt without the use of ML was made by Panayi et al. [24], for *Agaricus bisporus* mushroom yield prediction. The authors used advanced regression techniques in order to capture the influence of environmental variables during the production to the final yield produced. A very innovative aspect of their developed model was the ability to uniformly handle timeseries data from the production with different lengths.

This review aims to collect and present the most important studies in controlled environments which used Machine Learning and state-of-the-art Deep Learning models. The studies were selected based on the relevance of the problems they tackled e.g., important environmental variables and yield prediction. Furthermore, cases of environmental factors control with focus on the sensor installations are presented. Our ultimate goal is to create a complete picture of the state-of-the-art uses of ML models in CEA, find the similarities and potential uses in the mushroom case.

The remainder of this article is organized as follows. Section 2 presents the most important studies on CEA that utilize ML models. In Section 3, details on the used models in the examined studies are provided. In Section 4 studies that focus more on the installation and control of the environment are presented. Finally, in Section 5 we discuss future considerations based on the examined studies.

2 Machine Learning applications in CEA

The following collection of studies is the result of queries made on Scopus. Some of the terms used were "machine learning controlled environment agriculture" with many variations e.g., "deep learning controlled environment agriculture", "machine learning greenhouse agriculture", "controlled environment/greenhouse yield prediction" etc. From the queries made more than 150 papers were examined thoroughly in order to select the most prominent regarding the algorithms used and the results produced.

When cultivating a crop, the main goal is to use the observed data to avoid conditions that could prove catastrophic for the cultivation and/or to act in time to improve even further the desired yield. With these notion in mind, we can define two main categories for the application of Machine Learning in CEA:

1. **Climatic Variables**
2. **Produced Yield**

Each cultivated crop can be influenced by many environmental variables. Usually though, the most important of them are selected and monitored throughout the cultivation. But simply monitoring the present status of the controlled environment is not enough. In order to sustain the optimal conditions, the system must be proactive and not reactive, as some plants are sensitive to even small fluctuations to their growing conditions. This makes necessary the prediction of the future values of the climatic variables inside the controlled environment.

The final yield is strongly dependent on the climatic conditions of an installation. Establishing a correlation between the growing conditions and the yield produce is a powerful tool in the hands of the producers.

Monitoring and predicting both categories can lead to even better results. The use of Machine Learning (ML) algorithms and models becomes crucial if we take into consideration the complexity of these problems. In the recent years, there have been many studies that implemented ML algorithms with the aim to increase the overall efficiency of CEA. In our work, we have selected the most prominent studies found in bibliography that integrate the use of ML algorithms to deal with the aforementioned categories. In Figure 3, the number of papers per year but also the ML algorithm that was used.

It is obvious that the research and experiments on using ML to improve CEA has started almost 25 years ago. However, the bulk of the research has been conducted in the last two years, as our technology and knowledge are expanding. A very interesting observation is that in the last two years, new and more complex ML algorithms are utilized to tackle the problems. Up until 2019, the models used were the Artificial Neural Network architecture and simple regression, Bayesian and decision models. This seems to change from 2020 as a turn to more Deep Learning architectures is made.

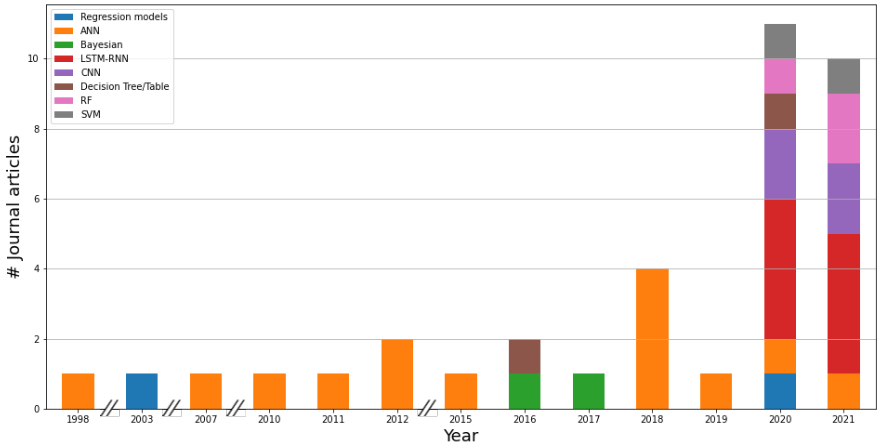

Fig. 3. # of studies per ML model and year in CEA bibliography.

Another interesting perspective is shown in Figure 4, where the total research papers per ML algorithm is shown.

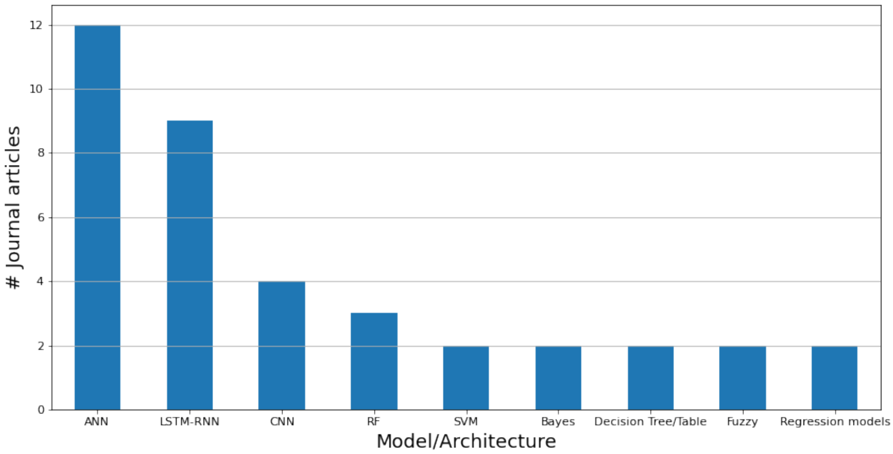

Fig. 4. # of papers per ML model in CEA bibliography.

We can see that the broad category of Neural Networks is dominating the preferences of the researchers. The most used models are the Artificial Neural Networks and their variations. However, combining the second and third places in the figure, the Deep Learning models surpass the ANNs. The preference for recurrent architectures has to do with the timeseries nature of the data used, as they can deal in a better way with them than the ANNs. The following chapter provides an overview of relevant research performed on the categories of environmental variables and yield prediction.

2.1 Climatic Variables Control and Prediction

The most important climatic variables in CEA are the air temperature, the air humidity and the CO_2 concentration. Light radiation is also very important in all plants but in the mushroom case it has a secondary role. A compact view of these variables can be seen in Table 1 In general, there has not been done any research on applying ML to light intensity.

Table 1. Table of ML research studies on CEA.

Variable	Model	Crop	Performance Metric			Author
			R^2	RMSE	MAE	
Air temperature (°C)	Regressive Models	Not defined	0.950	-	-	[25]
Air temperature(°C) - Relative humidity (%)	ANN	Not defined	0.977	1.81°C, 5.04%	-	[26]
Enthalpy(kJ/kg) - (Air temperature + Relative humidity)	ANN	Tomato	0.991	0.61 kJ/kg	-	[27]
Air temperature(°C) - Relative humidity (%)	ANFIS[1] (ANN)	Tomato	0.975	1.19°C, 2.16%	-	[28]
Air temperature(°C) - Relative humidity (%)	ANFIS[1] (ANN)	Stevia's leaves	0.98	1.63°C	1.18°C	[29]
Air temperature(°C) - Soil temperature(°C)	Dynamic Model	No crop	-	5.3°C, 3.45%	-	[30]
Air temperature(°C) - Relative humidity (%)	LM-RBF (ANN)	No crop	-	$9.99 * 10^{-6}$°C, 9.91*10^{-6}%	-	[31]
Air temperature(°C) - Yield prediction(Weight g, Leaf area m^3)		Banana	-	0.5294g, 0.4583m^3	0.466g, 0.2647m^3	[32]
Air temperature(°C) - Absolute humidity	LSTM	No crop	0.998, 0.9979	0.11°C, 0.40%	0.16°C, 0.62%	[33]
Air temperature(°C) - Relative humidity (%)	ERNN (ANN)	No crop	-	0.147	0.102	[34]
Air/Soil temperature(°C) - Relative Air/Soil humidity (%) - CO_2 concentration (ppm) - Illumination ($\mu mol/m^2/s$)	LSTM	tomato, cucumber, pepper	-	<0.6°C, <2%, - <5ppm, <160$\mu mol/m^2/s$	-	[35]
CO_2 concentration(ppm) - Air temperature (K)	ANN	No crop	0.97, 0.86	163ppm, 1.2K	-	[36]
CO_2 concentration($\mu mol/mol$)	ANN	No crop	0.97	19.9	-	[37]
CO_2 concentration($\mu mol/mol$)	LSTM	Irwin mango	0.78	-	-	[38]
Yield prediction (Kg/m^2)	ANN	Tomato	0.70	-	-	[39]
Yield prediction	ANN + GA[2]	Tomato	-	-	-	[40]
Yield prediction (Kg/m^2)	Dynamic ANN	Tomato	0.99	-	0.0697	[41]
Yield prediction (kg/10a) - Plant Growth (plant height, # leaves, leaf len./width, crown diam.)	Regressive Models	Strawberry	0.99,>0.8	101.55, [0.2, 1.26]	-	[42]
Yield prediction - Plant Growth (mm/day)	LSTM	Tomato, Ficus	-	0.047, 0.042(mm/day)	0.03, 0.03(mm/day)	[43]
Yield prediction (g/m^2)	LSTM - TCN[3]	Tomato	-	0.0026	0.0017	[44]
Plant Growth (stem diameter variations)	WT-ED-LSTM-AM[4]	Ficus	-	6.76	-	[45]

[1]Adaptive Neuro-Fuzzy Inference System. [2]Genetic Algorithm. [3]Temporal Convolutional Network.
[4]Wavelet Transform (WV) Encoder-Decoder (ED) Attention Mechanism (AM).

Variable	Model	Crop	Performance Metric			Author
			R^2	RMSE	MAE	
Water temperature (°C)	Decision Trees	water spinach	0.92847	0.0673	-	[46]
Yield prediction (oz)	Deep ANN, RF, SVR, Deep ANN+RF ensemble	garlic chives, basil, red chard, rainbow chard, arugula mint	0.81 (ensemble)	0.316 (DANN)	0.18 (DANN)	[47]
Plant Growth (image greenness)	CNN	green plants	-	-	-	[48]
Plant Health (image sickness area)	CNN	lettuce	-	-	-	[49]
Yield Prediction (grams)	RF, SVM, LSTM, CNN	garlic chives, basil, red chard, rainbow chard, arugula mint	0.752 (CNN)	7.9 (CNN)	4.537 (CNN)	[3]

Furthermore, in Table 1 we can see the models each study used and which was the crop under examination. For an easy comparison the three most used Performance Metrics are presented, however, each study might also used even more parameters to evaluate its results. The table, also, indicates the vast preference on the Neural Networks in its vanilla form as well as in custom implementations. The same can be said for Deep Learning implementations.

Temperature/Humidity Prediction Two of the major climatic factors that influence the crop are the temperature and the humidity. They have a strong influence in all the stages of its production and if the both of them rise or fall below an optimal range then the plant cannot develop properly and the crop will not reach the maximum yield potential. Because both of them are influenced by each other and by other existing climatic factors, the developed predictive models have as inputs various climatic variables of the controlled environment as well as of the natural environment outside the installation.

One of the earliest studies to predict the air temperature was conducted by Frausto et al. [25]. For the temperature prediction inside a unheated, naturally ventilated greenhouse, linear auto regressive models were used (ARX,ARMAX). The models had as inputs variables from outside the installation namely air temperature, air relative humidity, global solar radiation and cloudiness of the sky. One of the conclusions of the study was that the linear regression models were able to predict the temperature, except for ventilation periods, due to the highly non-linear behavior of the system.

Salazar et al. [26] turned their attention to Artificial Neural Networks (ANNs) to predict the air temperature as well as the humidity, because ANNs can deal with the non-linear relations existing between the climatic variables. In this study three models were made, two for the separate estimation of the variables and one for their combined calculation. All of the models outperformed related works with linear auto regressive models.

Salazar et al. [27] continued their work and implemented an approach to calculate temperature and humidity simultaneously through enthalpy, a variable which provides a combined description of these two variables. For its calculation an ANN model was selected with input variables of temperature, transpiration, ventilation and heating system was selected.

López-Cruz and Hernández-Larragoiti [28] implemented Neuro-Fuzzy models to predict air temperature and humidity. Even though humidity is a non-linear process the resulting model was simpler than that of the temperature. The authors comment that despite the good results obtained, further investigations must be conducted to decide if Neuro-Fuzzy models can be used for modelling and controlling variables more directly connected to the crops cultivated inside the installation.

Guerrero-Santana et al. [29] used Neuro-Fuzzy models to predict independently from each other the inside temperature and relative humidity. They implemented models with different input combinations and concluded that the models which included past values of the predicting variable produced better results from those which didn't use them.

Mohammadi et al. [30] used first order differential equations which are derived from energy balances of soil surface, inside air and roof cover. With this approach it was possible to model the heat transfer between elements of the installation and predict the desired variables.

Yue et al. [31] experimented with the RBF neural network in order to predict the variables of temperature and relative humidity inside the installation. The architecture of the RBF network provide more robustness in noisy data compared with the ANN. Moreover the use of the Levenberg-Marquardt algorithm during the training deals with the problem of local minimums and over-parameterization problems. This leads to the construction of an optimized network that produced goof results in the prediction of the environmental variables under examination

Soundiran et al. [32] used ANN to predict the inside temperature of the installation and crop yield. Regarding the temperature he started with a big fully connected ANN and applied the algorithm Optimal Brain Surgeon (OBS) [50] in order to prune the network and find the optimal neural network topology that could generalize best. ANN was also used to predict dry weight and leaf area of the crop, which are used to calculate the Net Assimilation Rate (NAR). NAR helps in quantitative terms to interpret crop yields under different environments. To tackle the problem of few training data in the yield prediction, the authors used the bootstrap ANN method to create a large number of samples from the original data.

Gharghory [33] tested the LSTM architecture for timeseries prediction of temperature and relative humidity. All the values are simulated from a greenhouse heating-cooling and ventilating (HCV) which takes into account the structure and properties of the installation, as well as outside weather variables as disturbances. The predicted values of the temperature and humidity are used as training and testing dataset for the selected LSTM network. The author concluded that the LSTM outperforms outperforms other conventional neural networks in the literature.

Aytenfsu et al. [34] explored the Elman Recurrent Neural Network (ERNN) as an option to predict the inside temperature and relative humidity of the installation based on data from outside. The difference of ERNN to the vanilla ANN is the internal feedback loop in the hidden layer that act as an internal state which saves the state of the hidden layer of the previous time unit. The design of the network is optimised and the results of the study shows that the selected networks is an efficient way to predict the target variables.

Liu et al. [35] explored the use of LSTM model for predicting six climate variables inside an greenhouse. The LSTM had as input the previous values of these variables and predicted their future values. In contrast with other studies which had one model per variable, this model produced the predictions for all variables at the same time. The high non-linearity of the problem at hand was dealt successfully by the LSTM. The authors compare their model with the RNN and GRU models. Both LSTM and GRU are variants of the RNN. The results showed that overall the LSTM had the best performance, which improved further when more data where used for the training of the model.

Taufiqurrahman et al. [46] tried to predict the future values of the water temperature of an aquaponic installation. Two regression models were explored namely, Adaboost Regressor and Decision Tree Regressor. The results of the study showed that the Adaboost Regressor, which is an ensemble model, produces better results and can predict more accurately the future values of the temperature, with regard to the MSE and R^2 metrics.

CO_2 Concentration Prediction The concentration of CO_2 inside a controlled environment is a very important factor for the plant growth. CO_2 is an essential component of photosynthesis and can be consider as a nutrient for the plants. When existing in optimal levels it can increase the plant growth and ultimately the productivity and yield of the installation. Monitoring, predicting and controlling the CO_2 concentration is crucial.

Linker et al. [36] conducted one of the earliest studies where ANN was used for modelling and predicting variables in a controlled environment. The authors made simplifications regarding the inputs and how they can affect the predicting variables in order to produce simple ANN models, due to the fact that available data was sparse in some parts of the input space. In this research it is shown that neural networks can be used instead of traditional descriptive models of a controlled environment.

Moon et al. [37] investigated the ANN as a solution to prediction the CO_2 concentration inside a greenhouse, when using multiple environmental variables. These variables were the major factors that affect the growth of a crop as well as the inside environment of an installation. The final estimations of the CO_2 concentration were good, showing that ANN is a valid approach for this problem. The authors noted that the ANN should be trained and test more with real data from other installations or from simulated data.

Moon et al. [38] expanded their previous work with ANNs and experimented with the LSTM architecture, utilizing the increasing deep learning trend on the domain. The model was fed with inputs of the inside and outside environments with the aim to predict the CO_2 concentration. The results showed an adequate performance and the CO_2 could be relatively well predicted. The author notes that for the improvement of the accuracy the time interval of 10 minutes for the data sampling is too long and should be reduced. Furthermore, they states that plant growth variables must be included as inputs to the model due to the fact that the plants act as a disturbance to the installation's environment that should be taken into account.

Finally, as has been described previously Liu et al. [35] have used the LSTM architecture to predict multiple climatic variables at the same time, including the CO_2 concentration of the installation.

2.2 Yield Prediction

The monitoring, predicting and controlling of the climatic factors inside a controlled environment is done with the higher purpose of achieving the best yield possible. An interesting problem is to correlate the environmental variables and their control with the yield. There have been many studies trying to predict the final yield from the climatic variables throughout the growing period of the crops.

Ehret et al. [39] utilized an ANN model in order to predict in weekly and daily basis the variables of yield, crop growth and water use in two greenhouses. Many models were created and tested with different input environmental parameters, depending on their Coefficient of determination R^2, in individual regression analysis wit the outputs. One of the authors' conclusions is that Radiation, Temperature and CO_2 Concentration have a strong influence on the output variables predicted.

Qaddoum et al. [40] in this work made a novel approach. The traditional ANN architecture is supplemented with a Genetic Algorithm for the training and the optimization of the model. The author showed that the simple model of ANN used up until then can be further improved with the used of specialized optimization algorithms and achieve better results in comparison with other NN models.

Salazar et al. [41] in this study used a DANN to predict the yield based on external conditions. The DANN is a modified version of a ANN with delays in its architecture that allows the model to use data and outputs of the past for the new predictions. Even though the best results are obtained when using all the selected variables as inputs, this study showed that the yield can be sufficiently predicted with the use of the external condition variable of solar radiation and the past yields produced.

Sim et al. [42] examined the relation between environmental variables and yield and plant growth through regression analysis. The results show that both yield and plant growth can be predicted well from environmental data. Moreover, the authors tested if the yield can be predicted from plant growth data and the final results indicated that the appropriate method to predict the yield is with the use of environmental data.

Alhnaity et al. [43] author deployed an LSTM in order to predict the yield and the plant growth of tomato and ficus respectively. The model's predictions were based on environmental timeseries data combined with past yield and plant growth. A comparison with other Machine Learning methods (Support Vector Regression and Random Forest) was implemented and it showed that the selected approach outperformed them.

Alhnaity et al. [45] proposed a novel approach to timeseries prediction. The framework built from the authors is comprised of four parts, each one contributing to the final prediction. The first part is comprised from a Wavelet Transform, which main use is to decompose the data and eliminate the noise. The second part is an Encoder-Decoder based on a LSTM network. This part is used to extract useful representation embeddings (feature extraction) which will be used next for multi-step prediction. The next part is comprised from a LSTM for the prediction. The outputs of the internal LSTM layers as well as its outputs are fed to the final part which is Attention Mechanisms. The final part is used to model a respective long-term dependence and decide the importance of each step of the data sequence. The previous four main parts of the architecture are followed by a single layer neural network which makes the final prediction. The model is used to predict intervals of 1, 6 and 12 hours. The results are compared with results from other state-of-the-art models (SVR, RF, MLP, GRU, LSTM) and are found to be better in all the examined metrics.

Gong et al. [44] combined the LSTM model with a TCN model in order to predict yield from past yield and environmental variables. The LSTM is used to extract representative features from the input sequence. The TCN further process the output of the previous part and extracts new features that are fed to a Fully Connected Network which makes the final prediction of yield. The best architecture for the network is explored and then it is compared to implementations of traditional machine learning and deep learning. The results indicate that the proposed method outperforms the all the other models examined. Finally, an important conclusion made from the study is that the historical yield is the most important factor for the future yield prediction.

Torres-Tello et al. [47] in this study experimented with three different Machine Learning/Deep Learning models, namely SVR, RF and DNN, on six different crops, inside an aeroponic installation. Also, the examined the case where the RF and DNN models are used as an ensemble model in order to further improve the results and tackle the problem of bad generalization due to overfitting. This study showed that a yield prediction model on this type of control environment is possible but the need for more data for the training are imperative.

The aim of Tenzer and Clifford [48] in this study was to determine the plant growth of the plants in a hydroponic environment based on the extend of their green parts which is an indicator for the biomass of the plant. The machine learning task falls into the category of image segmentation and object detection. The authors used in the form of Encoder-Decoder state-of-the-art CNN architectures in order to find the green areas in the image. One strength of the proposed method is that the models can find more green areas inside the images even though they are not labelled as green during the labelling. Another interesting finding was that the method was greatly color-agnostic, meaning that the models were able to identify leaves of the plants that are not green due to reflection (white) or bad health (brown), meaning that a more complete image of the plant can be detected. Even though the available dataset consisted of timeseries, the authors wanted to design a method applicable to non-timeseries data.

Pratame et al. [49] examined two state-of-the-art architectures in order to address the problem of unhealthy plant detection in hydroponic farms. The images obtained where manually labelled and fed into the models with different train/validation/test splits. The results showed that generally the YOLO architecture outperforms the Faster R-CNN with Inception V2. The overall good results from both architectures signify that this is a good approach to tackle this problem.

Torres-Tello and Ko [3] expanded their previous work [47], examining the more complex architectures of LSTM and CNN and expanding the variables used as inputs. However, the interesting contribution of this study is the attempt to interpret the features extracted from the models and how they affect the final prediction. To achieve this, SHapley Additive exPlanations (SHAP) [51] was used, identifying the input variables of CO_2 concentration and Water Temperature to be the most important.

3 Machine Learning Models

In this section, the models that are used in the studies presented in the previous section are examined. These models include the most important cases that are state-of-the-art and some simpler cases that were used when ML was starting to be utilized in CEA. With the aim to create a more complete understanding of their characteristic and their use, brief but concrete descriptions are made.

3.1 Regression Models

Regression models are supervised learning models, that utilize the values of input variables in order to produce a prediction/forecast value for the output variable. The model tries to find the relationship between the input and output variables. The most simple and common model of regression is the linear regression model [52], which assumes a linear relationship between input and output variables. There are more complex regression models e.g., ordinary least squares regression [53], which tries to find the hyperplane that minimizes the sum of squared differences between the observed data and the hyperplane.

Other important models are the family of AutoRegressive – Moving Average models (AR-MA) [54]. The two parts of the model (AR-MA) are respectively, regressing the lagged input variables and modelling the produced error as a linear combination of past errors. The family of these models includes extended versions of the model, where components are added to deal with specific problems, e.g. the AR-I-MA model [55] (”I” stands for Integrated) is used for non-stationary input data and the S-AR-I-MA model [56] (”S” stands for Seasonal) is used for input data that have a seasonal trends

3.2 Decision Trees/Table

Decision Trees (DT) are a predictive model that can be used both in classification and regression problems [57]. Input data are progressively divided into subgroups by comparing the data values. These divisions create a tree-like graph, which has nodes and branches. The branches represent the results of the comparisons while the nodes are of two types. The internal nodes are the comparisons while the final nodes (leaf nodes) are the final output of the tree (decision). The first node of the tree is also called a root.

Input values are being compared against the value of the internal nodes and the result of the comparison points which direction in the tree will be followed until a Decision is reached. A path from the root to a leaf node is considered a rule that led to the final decision. If the DT is shallow (few internal levels) and relatively small, then it can be represented as a Decision Table. In general, all Decision Tables can be represented as a Decision Tree, but the opposite is not always feasible.

3.3 Fuzzy Logic

Fuzzy logic [58] is an extension of the Boolean logic, which aims to depict everything not in a binary system of True or False, but integrating concepts of partially True/False. In mathematics, if True and False are respectively 1 and 0, then with Fuzzy logic the value of a variable can be any real number between 0 and 1.

This logic can be more useful than the binary logic, creating more complex models. An interesting example is the Adaptive Network-based Fuzzy Inference System (ANFIS) [59]. This system integrates principles from Fuzzy logic as well as from Neural Networks. It can approximate nonlinear functions through the inference systems of a set of fuzzy If-Then rules.

3.4 Bayesian Network

Bayesian Networks are probabilistic graphical models which analyze the data within the context of Bayesian inference [60]. They represent a set of variables and their conditional dependencies (and independence) as a Directed Acyclic Graph (DAG). Both problems of classification and regression can be solved by these models.

3.5 Random Forest

Random Forest algorithm [61] is an Ensemble Learning model for both classification and regression. More specifically, it combines instances of simpler/weaker learners in order to supplement their predictive performance. The weaker learners used by Random Forest are Decision Trees. If the task at hand is a classification task, then the model will output the class selected by the majority of the weaker learners. On the other hand, in a regression task, the model will output the average output of all the weak learners.

This algorithm has a lot of hyperparameters to be tuned, depending on each problem e.g., the number of weak estimators to be used or the depth of each estimator. Moreover, Random Forest is a widely preferred algorithm as it is a very robust model regarding overfitting but tend to lower performances when the problem becomes too complicated.

3.6 Support Vector Machine (SVM)

One of the most robust supervised machine learning models is the Support Vector Machines [62]. Given a set of input data, the algorithm tries to create the hyperlane that separates best the input classes. An improvement to the traditional algorithm is the introduction of kernels, which transform the input data to a higher dimension. With this "kernel trick", as it is called, the algorithm can find more easily the separating hyperlane, when the original data are non-linearly separatable.

The original algorithm was created to solve classification problems of two output classes. However, through the years the algorithm was modified in order to tackle multiclass classification, regression (SVR) [63] and clustering problems [64].

3.7 Artificial Neural Networks (ANN)

Artificial neural networks are computing models that are inspired by the way the neurons inside our brain work [65]. Both the biological and artificial neurons take signals as inputs, process them and produce an output that is fed as an input to other neurons.

Our brain consists of billions of neurons that are connected in a complex way, communicating with each other vigorously, enabling us to perform many complex functionalities in our life. In a similar way, the neurons of a network are connected to each other. However, we cannot fully imitate the many and arbitrary connections of the brain with a ANN, but we create a structured architecture which is comprised of layers [66]. Each layer has a number of neurons that have as inputs the outputs of the previous layer and its outputs are used as inputs from the next layer. Generally, there are three types of layers:

- **Input layer**: This layer receives the external input data.
- **Hidden layer(s)**: These layers are the powerhouse of the network, as it is in them that the learning happens.
- **Output layer**: This layer produces the final answer of the network to the problem, namely a prediction or a decision depending on the type of the problem.

The number of hidden layers separates the Neural Networks to the traditional Artificial Neural Networks (ANN), with one hidden layer and to the deep Artificial Neural Networks (DNN) with more than one hidden layer.

Both classification and regression problems can be solved by ANNs. Over the years, many research has been done on ANNs ranging from learning algorithms such as the popular perceptron algorithm [67] and back-propagation [68], to the aforementioned Adaptive Network-based Fuzzy Inference System (ANFIS) [59], Extreme Learning Machine (ELM) [69] and Radia Basis Function Networks [70].

Although the research for the ANNs begins in the 1940s, the DNNs are a relatively new area of Machine Learning which was unlocked with the progress of the hardware capabilities. Their big depth is able to learn more complex representations than the simple Machine Learning models. But this advantage comes at a cost as these deep networks can have massive architectures and their training can be very computationally demanding, while at the same time a great amount of training data is required. Nevertheless, right now the state-of-the-art performance in many domains is achieved by Deep Networks.

DNNs are not the only architecture found in Deep Learning. A neural network with fully connected layers is not always the best solution to specific problems. For this reason, other architectures utilize different mathematical ideas to best solve these specific problems. For example, the mathematical operation of convolution is used in Convolutional Neural Networks (CNN) to deal with image problems and the recurrent connections inside the architecture of a Recurrent Neural Network are exploited to deal with temporal problems. These networks will be briefly presented in the next paragraphs. A very descriptive presentation of Deep Neural Networks and their types is given in the book Deep Learning by Goodfellow et al. [71].

3.8 Recurrent Neural Networks (RNN)

A very interesting category of ANNs is the Recurrent Neural Networks (RNNs) [72],[73]. It utilizes its internal state of weights as a memory, making it ideal to deal with data sequences as inputs. The memory of the network helps it to make decisions and predictions for current or future steps by looking what has already happened.

In Figure 5 a general architecture of an RNN is shown. The initial internal state prior to training is the first green box named $a^{<0>}$. With **t** as the timestep, the other green boxes with the symbols $x^{<t>}$ are the inputs in each step, while the red boxes with the symbols $y^{<t>}$ are the outputs of each step. Inside the blue boxes the internal state $a^{<t>}$ is updated in each step. When we unroll the network, we can see it taking the form of a Directed Acyclic Graph. This depiction shows like the internal state is forwarded to the next time step for the calculations to take

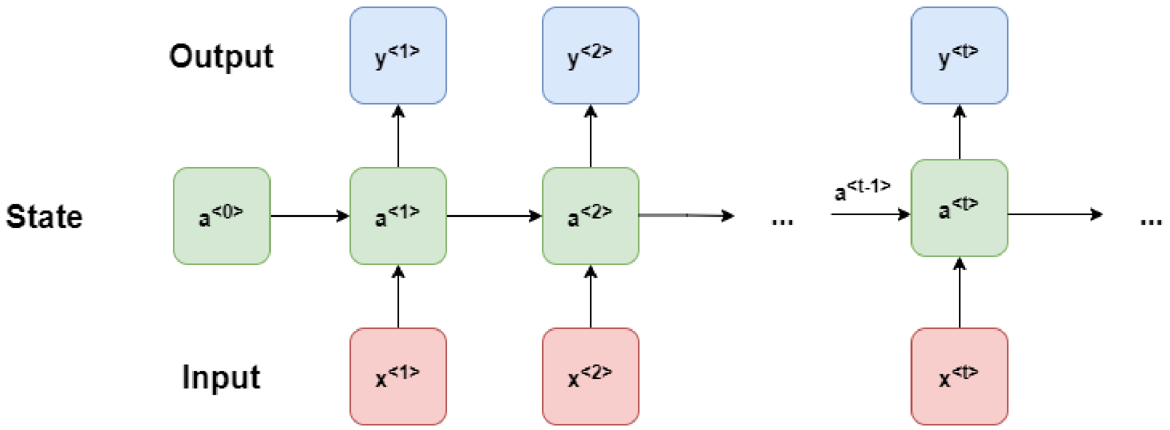

Fig. 5. Example of Recurrent Neural Network architecture.

place. Of course, as a neural network, RNNs can have many layers, meaning that the $y^{<t>}$ are used from the next layer as inputs. An example of input data for the RNN architecture is a timeseries of Temperature where $x^{<t>}$ can be a time sample from the timeseries at a specific moment.

Even though RNN models are excellent on paper for data sequences, in practice they are restricted by a phenomenon called "Vanishing/Exploding Gradient problem". This is a problem encountered when a vanilla RNN is trained and its size is very big. In the training of a vanilla RNN, the update of the network's weights begins from the final part of the networks and is propagated up to the initial part following the back-propagation algorithm. However, due to the many calculated derivatives, it is possible for the gradient values to progressively get smaller or larger. In the first case, the initial parts of the networks are not changed at all while in the second case very large updates happen causing destabilizing the network. Overall, the network cannot capture the long-term dependencies.

There is a number of practices that can alleviate the consequences of this problem, using proper weight initialization, network modifications and data pre-processing, but the problem has been solved efficiently by a new updated version of RNN named Long Short-Term Memory (LSTM).

3.9 Long Short-Term Memory (LSTM)

As a solution to the Vanishing/Exploding gradient problem of the RNN, the LSTM model was proposed by Hochreiter and Schmidhuber [74]. A cell of an LSTM networks is comprised of two parts, the long term and the short term (state) memory, which are essentially the flow of information of the entire network. The control of this flow happens from three parts (gates) inside an LSTM cell. The flow of information and its control is depicted as a black box in figure 6.

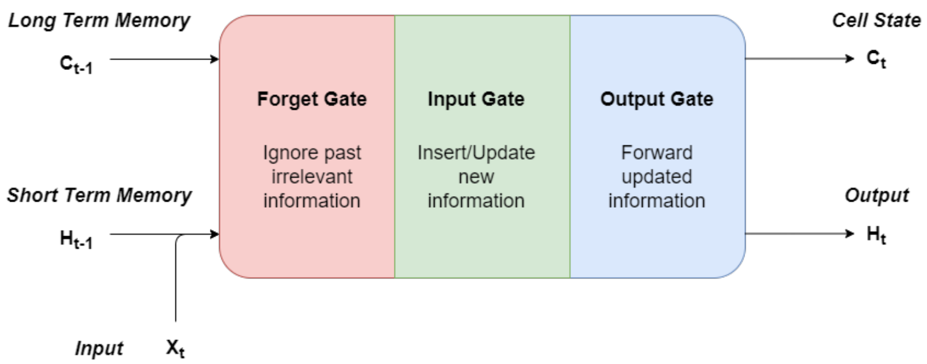

Fig. 6. LSTM cell information structure as black box.

The depicted gates and their roles that are visible in the image above are:

– **Forget Gate**, responsible for how much of the long term memory will be used in the calculations for the current state and passed on the next step.
– **Input Gate**, decides how much the input will influence the calculations for the current state.
– **Output Gate**, controls how much of the current state will be forwarded as an output of the cell (in the next step or as a result).

In figure 7 we can shed some light inside the structure of the LSTM cell and see where exactly are the control gates.

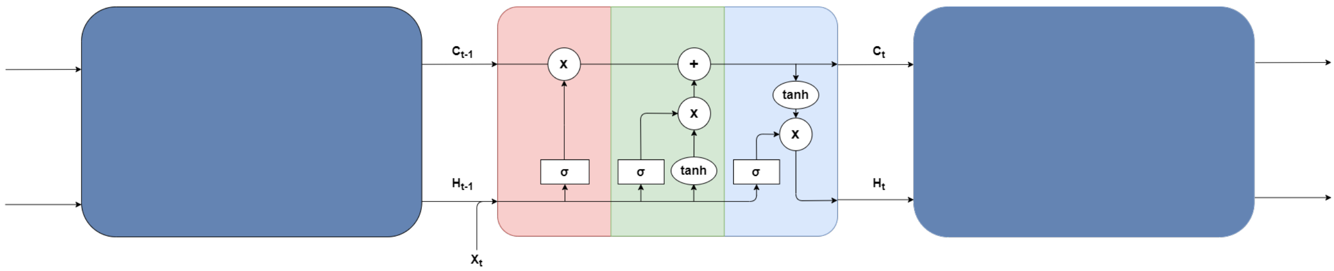

Fig. 7. More detailed LSTM cell structure.

The symbols existing in the above figure are σ for the sigmoid function, $\boldsymbol{x}$ for the multiplication operation, + for the addition operation and ***tanh*** for the hyperbolic tangent function. The $C_{<.>}$ and $H_{<.>}$ are used for the memory and output respectively. An example for input data for this model is a timeseries of Relative Humidity. Each time sample of the timeseries can be regraded as the X_t input of the figure.

3.10 Convolutional Neural Networks (CNN)

Convolutional Neural Networks is a type of Deep Neural Networks that has convolution as its main mathematical operation instead of the simple matrix multiplication. Even though they have appeared in late 90s [75], they have gained popularity relatively recently. These types of networks are usually used for classification and detection problems regarding images. They have the ability to handle input data with minor pre-processing, extracting in an unsupervised way features from the data. At the end of the CNN, usually a classifier is attached to utilize the extracted features, with the most popular option being an ANN.

With respect to the dimensions used in a CNN, three main categories exist:

– **1-D CNN**: Used to process sequential data, e.g., timeseries of Temperature, CO2 concentration.
– **2-D CNN**: Used to process images, e.g., RGB images of plants.
– **3-D CNN**: Used to process sequence of images, e.g., hyperspectral images of crops.

The architecture of a CNN is comprised from many different layers, that can be changed and modified to deal with the problem at hand. These layers usually are pooling layers, activation layers, fully-connected layers and of course, convolutional layers.

The CNN architecture is described extensively from Goodfellow [71], but a brief description of these basic layers is presented below:

– **Convolutional layer**: Applies the mathematical operation of convolution to the data.
– **Pooling layer**: Used to decrease the size of the convolution results by applying a max or average function on groups of elements. This happens in order to avoid great computational costs for the network and to aggregate information.
– **Activation layer**: This layer is used to control which information will be passed to next layers and which will be neglected. It introduces non-linearity in the network and helps to approximate the complex relationships between the variables of the system.
– **Fully-connected layer**: Usually an ANN, which serves the purpose of combining the features extracted from the network in order to make the classification.

An interesting trend is the use of 1-D CNNs, also called Temporal Convolutional Networks (TCNs), as an alternative to RNNs-LSTMs in various problems. Bai et al. [76] make an excellent and detailed description of a TCN

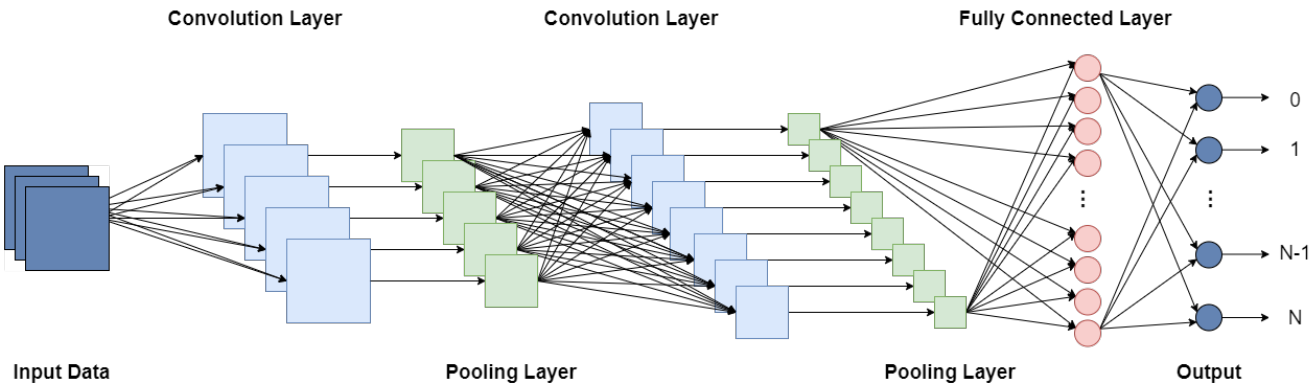

Fig. 8. A simple example of a 2D-CNN architecture.

architecture, comparing it in a number of sequence problems with LSTM architecture. The results show that TCN can deal effectively with the problem of this domain. The conclusion states, even though TCNs are overshadowed by the more popular RNN-based architectures, they should be regarded as a powerful tool for sequence modelling. This makes them an interesting and viable option in tasks dealing with timeseries.

4 IoT and Control installations in CEA

As its name states, the one of the biggest advantages of the CEA is the ability to monitor and control the conditions inside the installation. This way the optimal conditions can be created and maintained, ensuring a good growth of the crop and subsequently a good yield.

Regarding the mushroom cultivation, the high sensitivity of the mycelium towards its environment, makes very important the existence of sensors to collect data and monitor the conditions inside the installation.

To the best of our knowledge, most of the research on controlled environment mushroom cultivation, up until now, is focused on the IoT installations as it can be seen from column "Installation/Model" in Table 2, where a more compact view of the presented work is shown. The aim of these installations is to collect data during the growth period and control the environmental variables through actuators, in order to keep the growing conditions optimal. Most of the time the processing of the collected data is happening without machine learning, in simple logic of "if-else" conditions, e.g. if the temperature rises above 25^{o}C, activate a cooling mechanism.

Moreover, in Table 2 the parameters that are being monitored in each study are considered the most important by the authors. These parameters are in the column "Parameters monitored".

Here, we present the most recent and important works on IoT control installations found in CE mushroom cultivation. One of the most important factors when following the IoT approach is to minimize the energy consumption while having a robust controlling installation.

Zhao and Zhu [77] made a study on CE of *Pleurotus Eryngii* mushrooms, creating a system that collected and transmitted wirelessly data. At the same time, a user could control the environmental conditions through actuators remotely, either in a manual or automatic way. The system was developed with the aim of saving energy, while improving the quality and yield of the mushrooms.

Yu and Xue [13] had the same target when conducting their study. The IoT installation that was developed inside a Shiitake Mushroom farm. The results showed that the monitoring and controlling of the farm with IoT were able to increase the yield and the quality of the mushrooms while optimizing the usage of the farm's resources.

The minimization of cost was the aim of Ariffin et al. [15] for an Oyster mushroom farm. In this study, the authors made an effort to develop an automatic IoT-based Climate Control system, that was designed with minimal cost in terms of hardware and farm resources used. The result was a practical system, that reduced needed human labor for monitoring and maintaining the environment of the installation.

An interesting way to control the actuators based on the collected data can be done using Fuzzy Logic. This was examined by Ardabilia et al. [78] where a comparison between a controller with Fuzzy Logic and a controller of digital ON/OFF is done. The results showed that the Fuzzy controller was able to perform better, with respect to the responses in the controlling process as well as in energy consumption.

Table 2. IoT and Control studies in CEA.

Parameters monitored	Installation/Model	Crop	Author
Air temperature, Humidity - Light intensity, CO_2 concentration	Hardware Installation	Pleurotus eryngii Mushroom	[77]
Air temperature, Humidity - CO_2 concentration	Fuzzy Logic	Not defined Mushroom species	[78]
Air temperature, Humidity	ELM[1]	Oyster Mushroom	[79]
Air temperature, Humidity	Fuzzy Logic	Oyster Mushroom	[80]
Air temperature, Humidity - CO_2 concentration	IoT installation	Shiitake Mushroom	[13]
Air temperature, Humidity	IoT installation	Oyster Mushroom	[15]
Air/Soil/Leaves temperature - Solar,PAR radiation, Humidity - Rainfall, Wind Velocity	Bayesian networks	Not defined	[12]
Air/Water temperature, Humidity - Light intensity, pH, EC[2]	Decision Table Rules	Not defined	[81]
Water temperature, Humidity - Light intensity, pH, EC[2]	Bayesian Network	Iceberg lettuce	[82]
Air temperature, Water level - Humidity, Light intensity, pH	Deep Neural Network (DNN)	Tomato	[83]
Water temperature, Humidity - Light intensity	Random Forest	Lettuce	[84]
Water temperature, Water Level - Humidity, Light intensity	Deep Neural Network (DNN)	Lettuce	[14]
Air/Floor/Ceiling/Wall temperature - Solar radiation, Air/Ground temperature	Model Predictive Control	Tomato	[85]

[1]Extreme Learning Machine. [2]Electro-Conductivity.

In the same direction, Amen and Villaverde [80] tried to show that automatic control is superior to human monitoring and controlling environmental variables of an installation. Their research concluded that automatic control system using Fuzzy Logic was able to regulate temperature and humidity better than manual control.

Fuady et al. [79] examined a machine learning approach for the controlling of the conditions inside an installation for Oyster mushrooms. In their study a comparison between Extreme Learning Machine (ELM) and ANN is implemented. Both of the models are able to deal with the problem at hand, but the results show that the ELM is able to reach stable conditions at the half time needed from the ANN.

Other notable studies on control in CEA but not on mushroom are presented below.

Peuchpanngarm et al. [81] designed a low-cost monitoring systems for hydroponic farming. The application keeps track of the input variables and acts accordingly, controlling the environment, when some manually written rules are satisfied. Even though this implementation is for DIY farms, there is potential to be used in small and medium farms, combining machine learning techniques on the data collected from the sensors.

Del Sagrado et al. [12] used Bayesian networks as their model in order to influence air temperature through ventilation aperture control. Bayesian models are not used for regression but for classification. Respecting that, the authors made all their variables discrete applying a supervised discretization algorithm [86] to obtain, based on the class, the optimal number of states of each variable to discretize. Then, the discrete variable of ventilation aperture is used as the class all the other variables are used as predicting variables.

Chen and You [85] used Model Predictive Control (MPC) in order to predict disturbances in the system and to control the temperature within imposed constrains with respect to the present crop.

Alipio et al. [82] utilized the data collected from sensors placed in a hydroponic environment to control it using Bayesian Networks. The generated networks perform the predictive analysis, feeding the actuators of the systems with decision for the environmental control. The results showed that the use of the Bayesian Networks improved the yield up to 60% compared to manual control.

Mehra et al.[83] in this study addressed the environmental control of a hydroponic installation as a classification problem, with 8 different output classes as actions.They selected a Deep Neural Network (DNN) as their model, which takes as inputs samples taken from sensors and decides what action should the actuators make. The model showed good accuracy and the plants that were produced under its control in the controlled environment had bigger height than the plants grown in soil.

Nugroho et al. [14] developed a custom hydroponic installation with an IoT system for controlling it. The decision for the control were supplemented by a DNN which used data from the installed sensors as inputs. The task was addressed as a classification problem with 8 different control actions (classes) as the output of the DNN. The authors conclude that the accuracy of the model was greater than the respective manual in controlling the environment

Karuniawati et al. [84] in this study, examined how the fusion of light intensity, humidity and temperature sensor data in combination with a Random Forest model can address the light control problem. The lights are controlled in a binary way, on or off. The authors conclude that the data fusion is better as it utilize multiple information inputs to produce a final decision.

5 Future considerations

The improvement of CEA is highly dependable on the improvement of the methods used for predicting and controlling the environment, in order to provide even better conditions for the cultivated plants. This can be done by constantly integrating new and better hardware, but also by improving and combining the algorithms used. This can lead to more sustainable strategies in production, while increasing the efficiency in crop production.

Regarding the algorithms, the trend shows that new algorithms being explored and this should be expanded in the future. The use of Deep Learning models is gaining ground and most probably will dominate the preferences of the researchers, due to the high amounts of data that can be collected from the installed sensor systems. Furthermore, this preference can be attributed to the recent advances in the domain of Artificial Intelligence, but also, to the technological advances in the need hardware to train these demanding models. Moreover, the retrieved algorithms from the literature, comprise an inventory of ML models in CEA act as a proof of concept as their results have shown that predicting and controlling the environment in CEA can be done in a very accurate and efficient way.

Even though many crops have been examined in the relevant studies, a lack of research exists on mushrooms, despite their importance in human diet and their market value. A complete study of the mushroom chain will surely be of great importance to the industry. A data-driven approach will be very interesting as it could examine the combination of the different topics the presented studies focus on. Apart from using state-of-the-art ML models to predict the climatic values and yield of mushrooms, an interesting direction is to use ML models to examine and create the correlation between the different experimental cultivation conditions and the final yield and quality of the product. Based on the extracted knowledge of such a study, a complete and robust prediction and control of the environment can be developed in order to reach and sustain optimal conditions throughout the cultivation period, subsequently leading to the best yield and quality. A study like that can pave the way for the development of a full data-driven ML model for mushroom cultivation in CEA. The selected models will be used as a basis to develop MUSHNOMICS algorithms, utilizing any relevant knowledge and procedures extracted for the collected studies that were presented.

References

1. José A. Aznar-Sánchez, Juan F. Velasco-Muñoz, Belén López-Felices, and Isabel M. Román-Sánchez. An analysis of global research trends on greenhouse technology: Towards a sustainable agriculture. *International Journal of Environmental Research and Public Health*, 17(2), 2020.
2. Mamta D. Sardare . A REVIEW ON PLANT WITHOUT SOIL - HYDROPONICS. *International Journal of Research in Engineering and Technology*, 02(03):299–304, March 2013.
3. Julio Torres-Tello and Seok-Bum Ko. Interpretability of artificial intelligence models that use data fusion to predict yield in aeroponics. *Journal of Ambient Intelligence and Humanized Computing*, Sep 2021.
4. Ranka Junge, Bettina König, Morris Villarroel, Tamas Komives, and M. Jijakli. Strategic points in aquaponics. *Water*, 9(3):182, March 2017.
5. A. De Gelder, J. A. Dieleman, G.P.A. Bot, and L. F. M. Marcelis. An overview of climate and crop yield in closed greenhouses. *The Journal of Horticultural Science and Biotechnology*, 87(3):193–202, January 2012.
6. Kurt Benke and Bruce Tomkins. Future food-production systems: vertical farming and controlled-environment agriculture. *Sustainability: Science, Practice and Policy*, 13(1):13–26, January 2017.
7. Nicholas Engler and Moncef Krarti. Review of energy efficiency in controlled environment agriculture. *Renewable and Sustainable Energy Reviews*, 141:110786, May 2021.
8. P. Srivani, Yamuna Devi C., and S.H. Manjula. A controlled environment agriculture with hydroponics: Variants, parameters, methodologies and challenges for smart farming. In *2019 Fifteenth International Conference on Information Processing (ICINPRO)*. IEEE, December 2019.
9. Guilherme Barbosa, Francisca Gadelha, Natalya Kublik, Alan Proctor, Lucas Reichelm, Emily Weissinger, Gregory Wohlleb, and Rolf Halden. Comparison of land, water, and energy requirements of lettuce grown using hydroponic vs. conventional agricultural methods. *International Journal of Environmental Research and Public Health*, 12(6):6879–6891, June 2015.
10. S. Ragaveena, A. Shirly Edward, and U. Surendran. Smart controlled environment agriculture methods: a holistic review. *Reviews in Environmental Science and Bio/Technology*, 20(4):887–913, Dec 2021.
11. Rakiba Rayhana, Gaozhi Xiao, and Zheng Liu. Internet of things empowered smart greenhouse farming. *IEEE Journal of Radio Frequency Identification*, 4(3):195–211, September 2020.
12. Bayesian networks for greenhouse temperature control. *Journal of Applied Logic*, 17:25–35, 2016. SOCO13.
13. Mohamed Rawidean Mohd Kassim, Ibrahim Mat, and Ismail Mat Yusoff. Applications of internet of things in mushroom farm management. In *2019 13th International Conference on Sensing Technology (ICST)*, pages 1–6, 2019.
14. Edi Dwi Nugroho, Aji Gautama Putrada, and Andrian Rakhmatsyah. Predictive control on lettuce nft-based hydroponic iot using deep neural network. In *2021 International Symposium on Electronics and Smart Devices (ISESD)*, pages 1–6, 2021.
15. Muhammad Azizi Mohd Ariffin, Muhammad Izzad Ramli, Mohd Nazrul Mohd Amin, Marina Ismail, Zarina Zainol, Nor Diana Ahmad, and Nursuriati Jamil. Automatic climate control for mushroom cultivation using iot approach. In *2020 IEEE 10th International Conference on System Engineering and Technology (ICSET)*, pages 123–128, 2020.
16. Tariqul Islam, Nasrul Hamidin, Zarina Zakaria, and Mohd Ishak. A competitive study on higher yield performance in indoor optimized environment and outdoor cultivation of pleurotus pulmonarius. *MAYFEB Journal of Agricultural Science*, 2:13–27, 11 2016.
17. Jegadeesh Raman, Kab-Yeul Jang, Youn-Lee Oh, Minji Oh, Ji-Hoon Im, Hariprasath Lakshmanan, and Vikineswary Sabaratnam. Cultivation and nutritional value of prominent ipleurotus/i spp.: An overview. *Mycobiology*, 49(1):1–14, November 2020.
18. Carmen Sánchez. Cultivation of pleurotus ostreatus and other edible mushrooms. *Applied Microbiology and Biotechnology*, 85(5):1321–1337, December 2009.
19. Marcelo Barba Bellettini, Fernanda Assumpção Fiorda, Helayne Aparecida Maieves, Gerson Lopes Teixeira, Suelen Ávila, Polyanna Silveira Hornung, Agenor Maccari Júnior, and Rosemary Hoffmann Ribani. Factors affecting mushroom pleurotus spp. *Saudi Journal of Biological Sciences*, 26(4):633–646, May 2019.
20. Sjaak Wolfert, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt. Big data in smart farming – a review. *Agricultural Systems*, 153:69–80, May 2017.
21. Konstantinos Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. Machine learning in agriculture: A review. *Sensors*, 18(8):2674, August 2018.
22. Andreas Kamilaris and Francesc X. Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90, April 2018.
23. Abigail Cohen, Gerry Chen, Eli Berger, Sushmita Warrier, Guanghui Lan, Emily Grubert, Frank Dellaert, and Yongsheng Chen. Dynamically controlled environment agriculture: Integrating machine learning and mechanistic and physiological models for sustainable food cultivation. *ACS EST Engineering*, XXXX, 11 2021.
24. E. Panayi, G.W. Peters, and G. Kyriakides. Statistical modelling for precision agriculture: A case study in optimal environmental schedules for agaricus bisporus production via variable domain functional regression. *PLoS ONE*, 12(9), 2017. cited By 5.
25. Modelling greenhouse temperature by means of auto regressive models. *Biosystems Engineering*, 84(2):147–157, 2003.
26. Raquel Salazar, Irineo López, and Abraham Rojano. A neural network model to control greenhouse environment. In *2007 Sixth Mexican International Conference on Artificial Intelligence, Special Session (MICAI)*, pages 311–318, 2007.

27. Raquel Salazar, Abraham Rojano, Irineo López, and Uwe Schmidt. A model for the combine description of the temperature and relative humidity regime in the greenhouse. In *2010 Ninth Mexican International Conference on Artificial Intelligence*, pages 113–117, 2010.
28. I.L. López-Cruz and L. Hernández-Larragoiti. Neuro-fuzzy models for air temperature and humidity of a greenhouse. *Acta Horticulturae*, 927:611–618, 2012. cited By 1.
29. A. Guerrero-Santana, I. López-Cruz, A. Ruiz-García, E. Fitz-Rodríguez, and R. Salazar-Moreno. Prediction of air temperature and relative humidity in a solar greenhouse dryer using neuro-fuzzy models. 2018. cited By 0.
30. B. Mohammadi, S.F. Ranjbar, and Y. Ajabshirchi. Application of dynamic model to predict some inside environment variables in a semi-solar greenhouse. *Information Processing in Agriculture*, 5(2):279–288, 2018. cited By 27.
31. Youjun Yue, Jun Quan, Hui Zhao, and Hongjun Wang. The prediction of greenhouse temperature and humidity based on lm-rbf network. In *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 1537–1541, 2018.
32. R. Soundiran, T.K. Radhakrishnan, and S. Natarajan. Modeling of greenhouse agro-ecosystem using optimally designed bootstrapping artificial neural network. *Neural Computing and Applications*, 31(11):7821–7836, 2019. cited By 4.
33. S.M. Gharghory. Deep network based on long short-term memory for time series prediction of microclimate data inside the greenhouse. *International Journal of Computational Intelligence and Applications*, 19(2), 2020. cited By 3.
34. Samuel A. Aytenfsu, Asrat M. Beyene, and Tameru H. Getaneh. Controlling the interior of greenhouses using elman recurrent neural network. In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, pages 32–39, 2020.
35. Yuwen Liu, Dejuan Li, Shaohua Wan, Fan Wang, Wanchun Dou, Xiaolong Xu, Shancang Li, Rui Ma, and Lianyong Qi. A long short-term memory-based model for greenhouse climate prediction. *International Journal of Intelligent Systems*, n/a(n/a).
36. Optimal co2 control in a greenhouse modeled with neural networks. *Computers and Electronics in Agriculture*, 19(3):289–310, 1998.
37. T.W. Moon, D.H. Jung, S.H. Chang, and J.E. Son. Estimation of greenhouse co2 concentration via an artificial neural network that uses environmental factors. *Horticulture Environment and Biotechnology*, 59(1):45–50, 2018. cited By 11.
38. T. Moon, H.Y. Choi, D.H. Jung, S.H. Chang, and J.E. Son. Prediction of co2 concentration via long short-term memory using environmental factors in greenhouses. *Horticultural Science and Technology*, 38(2):201–209, 2020. cited By 2.
39. Neural network modeling of greenhouse tomato yield, growth and water use from automated crop monitoring data. *Computers and Electronics in Agriculture*, 79(1):82–89, 2011.
40. K. Qaddoum, E. Hines, and D. Iliescu. Yield prediction technique using hybrid adaptive neural genetic network. *International Journal of Computational Intelligence and Applications*, 11(4), 2012. cited By 3.
41. R. Salazar, I. López, A. Rojano, U. Schmidt, and D. Dannehl. Tomato yield prediction in a semi-closed greenhouse. *Acta Horticulturae*, 1107:263–269, 2015. cited By 8.
42. H.S. Sim, D.S. Kim, M.G. Ahn, S.R. Ahn, and S.K. Kim. Prediction of strawberry growth and fruit yield based on environmental and growth data in a greenhouse for soil cultivation with applied autonomous facilities. *Horticultural Science and Technology*, 38(6):840–849, 2020. cited By 3.
43. B. Alhnaity, S. Pearson, G. Leontidis, and S. Kollias. Using deep learning to predict plant growth and yield in greenhouse environments. *Acta Horticulturae*, 1296:425–431, 2020. cited By 1.
44. Liyun Gong, Miao Yu, Shouyong Jiang, Vassilis Cutsuridis, and Simon Pearson. Deep learning based prediction on greenhouse crop yield combined tcn and rnn. *Sensors*, 21(13), 2021.
45. An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth. *Information Sciences*, 560:35–50, 2021.
46. Akbar Taufiqurrahman, Aji Gautama Putrada, and Febri Dawani. Decision tree regression with adaboost ensemble learning for water temperature forecasting in aquaponic ecosystem. In *2020 6th International Conference on Interactive Digital Media (ICIDM)*, pages 1–5, 2020.
47. Julio Torres-Tello, Suganthi Venkatachalam, Lyman Moreno, and Seok-Bum Ko. Ensemble learning for improving generalization in aeroponics yield prediction. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2020.
48. Mark L Tenzer and Nicholas C Clifford. A digital green thumb: Neural networks to monitor hydroponic plant growth. In *2020 Systems and Information Engineering Design Symposium (SIEDS)*, pages 1–6, 2020.
49. Ichsan Yudha Pratama, Abdi Wahab, and Mudrik Alaydrus. Deep learning for assessing unhealthy lettuce hydroponic using convolutional neural network based on faster r-cnn with inception v2. In *2020 Fifth International Conference on Informatics and Computing (ICIC)*, pages 1–6, 2020.
50. Babak Hassibi and David G. Stork. Second order derivatives for network pruning: Optimal brain surgeon. In *NIPS*, 1992.
51. Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *ArXiv*, abs/1705.07874, 2017.
52. A. Schneider, G. Hommel, and M. Blettner. Linear regression analysis - part 14 of a series on evaluation of scientific publications. *Deutsches Arzteblatt*, 107(44):776–782, 2010. Cited By :223.
53. G. D. Hutcheson. Chapter 3 — ordinary least-squares regression. In *The multivariate social scientist*, pages 56–113, 1999.
54. P. A. P. Moran and Peter Whittle. Hypothesis testing in time series analysis. 1951.
55. S.L. Ho and M. Xie. The use of arima models for reliability forecasting and analysis. *Computers and Industrial Engineering*, 35(1-2):213–216, 1998. cited By 176.

56. Adhistya Permanasari, Indriana Hidayah, and Isna Alfi Bustoni. Sarima (seasonal arima) implementation on time series to forecast the number of malaria incidence. pages 203–207, 10 2013.
57. Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and C. J. Stone. Classification and regression trees. 1983.
58. Vilém Novák, Irina Perfilieva, and Jiri Mockor. Mathematical principles of fuzzy logic. 1999.
59. J.-S.R. Jang. Anfis: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3):665–685, 1993.
60. Finn V. Jensen and Thomas D. Nielsen. *Bayesian Networks and Decision Graphs.* Springer Publishing Company, Incorporated, 2nd edition, 2007.
61. L Breiman. Random forests. *Machine Learning*, 45:5–32, 10 2001.
62. C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. cited By 31418.
63. Debasish Basak, Srimanta Pal, and Dipak Patranabis. Support vector regression. *Neural Information Processing – Letters and Reviews*, 11, 11 2007.
64. Asa Ben-Hur, David Horn, Hava Siegelmann, and Vladimir Vapnik. Support vector clustering. *Journal of Machine Learning Research*, 2:125–137, 11 2001.
65. Warren Mcculloch and Walter Pitts. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:127–147, 1943.
66. Simon S. Haykin. *Neural networks and learning machines.* Pearson Education, Upper Saddle River, NJ, third edition, 2009.
67. F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.
68. David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning Representations by Back-propagating Errors. *Nature*, (6088):533–536.
69. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1):489–501, 2006. Neural Networks.
70. David Broomhead and David Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. *ROYAL SIGNALS AND RADAR ESTABLISHMENT MALVERN (UNITED KINGDOM)*, RSRE-MEMO-4148, 03 1988.
71. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning.* MIT Press, 2016. http://www.deeplearningbook.org.
72. David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning internal representations by error propagation. 1986.
73. M. I. Jordan. Serial order: A parallel distributed processing approach. Technical Report ICS Report 8604, Institute for Cognitive Science, University of California, San Diego, 1986.
74. Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997.
75. Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. Object recognition with gradient-based learning. In *Shape, Contour and Grouping in Computer Vision*, page 319, Berlin, Heidelberg, 1999. Springer-Verlag.
76. Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *ArXiv*, abs/1803.01271, 2018.
77. Li Zhao and Xuejun Zhu. The development of remote monitoring system for cultivation environment of pleurotus eryngii. In *2015 IEEE International Conference on Information and Automation*, pages 2643–2648, 2015.
78. Modeling and comparison of fuzzy and on/off controller in a mushroom growing hall. *Measurement*, 90:127–134, 2016.
79. G. M. Fuady, A.H. Turoobi, M. N. Majdi, M. Syaiin, R.Y. Adhitya, Isa Rachman, F. Rachman, M. A. P. Negara, A. Soeprijanto, and R.T. Soelistijono. Extreme learning machine and back propagation neural network comparison for temperature and humidity control of oyster mushroom based on microcontroller. In *2017 International Symposium on Electronics and Smart Devices (ISESD)*, pages 46–50, 2017.
80. Jennifer dela Cruz-del Amen and Jocelyn Flores Villaverde. Fuzzy logic-based controlled environment for the production of oyster mushroom. In *2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, pages 1–5, 2019.
81. Chanya Peuchpanngarm, Pantita Srinitiworawong, Wannisa Samerjai, and Thanwadee Sunetnanta. Diy sensor-based automatic control mobile application for hydroponics. In *2016 Fifth ICT International Student Project Conference (ICT-ISPC)*, pages 57–60, 2016.
82. Melchizedek I. Alipio, Allen Earl M. Dela Cruz, Jess David A. Doria, and Rowena Maria S. Fruto. A smart hydroponics farming system using exact inference in bayesian network. In *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, pages 1–5, 2017.
83. Iot based hydroponics system using deep neural networks. *Computers and Electronics in Agriculture*, 155:473–486, 2018.
84. Septinia Karuniawati, Aji Gautama Putrada, and Andrian Rakhmatsyah. Optimization of grow lights control in iot-based aeroponic systems with sensor fusion and random forest classification. In *2021 International Symposium on Electronics and Smart Devices (ISESD)*, pages 1–6, 2021.
85. Wei-Han Chen and Fengqi You. Efficient greenhouse temperature control with data-driven robust model predictive. In *2020 American Control Conference (ACC)*, pages 1986–1991, 2020.
86. Usama M. Fayyad and Keki B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *IJCAI*, 1993.