

RESEARCH

Open Access



A deep learning model for predicting risks of crop pests and diseases from sequential environmental data

Sangyeon Lee¹ and Choa Mun Yun^{2*}

Abstract

Crop pests reduce productivity, so managing them through early detection and prevention is essential. Data from various modalities are being used to predict crop diseases by applying machine learning methodology. In particular, because growth environment data is relatively easy to obtain, many attempts are made to predict pests and diseases using it. In this paper, we propose a model that predicts diseases through previous growth environment information of crops, including air temperature, relative humidity, dew point, and CO₂ concentration, using deep learning techniques. Using large-scale public data on crops of strawberry, pepper, grape, tomato, and paprika, we showed the model can predict the risk score of crop pests and diseases. It showed high predictive performance with an average AUROC of 0.917, and based on the predicted results, it can help prevent pests or post-processing. This environmental data-based crop disease prediction model and learning framework are expected to be universally applicable to various facilities and crops for disease/pest prevention.

Keywords Crop disease, Pest, Deep learning, Environmental data, Prevention

Background

Crop diseases and pests increase the cost of production by reducing the yield and the quality of crops and increase the required labor, also cause environmental damage due to the overuse of chemicals to treat pests [1]. They are one of the major factors threatening food security, which causes about 15–40% of global grain yield loss [2–4]. In particular, powdery mildew, a significant pest and disease of strawberries covered in this paper, results in up to 70% yield loss [5, 6]. Also, gray mold, another major condition, generally damages 15–20% of strawberries and more than

50% in severe cases [7]. Pesticides used to prevent this are also used in trillions of US dollars, of which only 0.1% kills the target, and the rest is absorbed and distributed into the environment and causes pollution [8, 9]. In order to minimize the economic and environmental damages caused by these pests, several attempts have been made to predict pests [1, 10]. These disease modeling attempts create several mathematical models that explain the occurrence and development of pests by various factors and use it to predict the occurrence and severity of pests [11, 12], understand causes and pests reduce the yield and quality of crops, increase costs due to increased labor for pest treatment, the development of disease [13–15]. It is used for assisting tactical decision making [16–18].

The major factors of pest modeling include the growth environment of crops, such as temperature, humidity, and precipitation. Among various factors that affect the occurrence of pests and diseases, the growing environment of crops is known to have a significant influence [1,

*Correspondence:

Choa Mun Yun
sangyeonlee230@gmail.com

¹ Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

² Sherpa Space Inc., Daejeon 34028, Republic of Korea



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

10, 19], and is often used for modeling because it is easier to collect and manage than other factors, such as host genomics [20], host nutrients [21], and microbial pathogen [22]. For example, A-scab [11], a model that predicts *Venturia inaequalis* infection in apples, calculates several intermediate variables using environmental variables of air temperature, rainfall, relative humidity, and leaf wetness and builds the criteria in the flow diagram. The presence or absence of pests is predicted corresponding to the flow diagram. INDIGOBLAST, a late blight prediction model of potatoes, also predicts pest occurrence by calculating the overall late blight severity index using thermal suitability and humidity suitability index calculated by temperature and humidity recordings [23].

These conceptual modeling methods have the advantage of being able to interpret predictions such as the underlying biological mechanism through [10] model. However, these models use many variables to predict pests and diseases, and a deep understanding of both target pests and crops is essential in integrating the variables. As a result, a new model must be built or adjust existing models heavily if the target crop or pest and disease for the model application are changed. However, creating models for an extensive range of crops and pests is impossible. The data-driven method has been actively studied because it can use large-scale and large-capacity complex data thanks to the recent development of big data and deep learning technology [24, 25]. Environment data [24, 26, 27] and image data [25, 28–30] are expanding to multi-modal [31]. Among various types of data, growing environment data is easy to collect on a large scale because the hardware for data collection is relatively easy to obtain and can be easily collected through devices installed in facility farms. To our best knowledge, there is still no model that predicts or forecasts the pest infestations of crops only with the growth environment information of crops in a universal situation by applying deep learning methodology.

In this paper, we propose a deep learning-based predictive modeling method applicable for preventing disease and pest infestation by predicting them through crop growth environmental data and calculating a linear risk score leading to a normal-infestation state. A prediction model was built with growth environment data collected from multiple facilities to show that the methodology is universally applicable regardless of the diversity of facility farm facilities, the diversity of crops, and target pests.

Results

For most of the crops and pests used in this study, normal and infested crops were sorted with the status in the latent space using the test dataset (Fig. 1 and Additional file 1: Fig. S1). Data points tend to be classified and

aligned from normal to infested clusters from coordinates $\alpha=0$ to $\alpha=1$. Due to the loss function used in the model, it was trained to gather normal points near $\alpha=0$ and the pest points near $\alpha=1$ on the latent space. The similarity of the cumulative environmental features is also represented by the distance between the points. In particular, there is a section in which normal and infested samples are mixed, representing the transition stage. It is the point that many samples are changing their status from normal to infested state. In the transition stage, the number of infested crops gradually increases with the change in the growth environment. It shows that it can be used as a model that predicts disease onset by increasing the risk score before the outbreak because it gives a continuous risk score rather than a binary classification result.

We tested how well the condition of the crop can be predicted from the accumulated environmental data using the risk score from the latent space. In all cases, the result shows that the risk score was significantly higher in the infested samples than in the normal cases ($p < 0.05$), with some overlaps in the transition stage (Fig. 2a). Normal and infested samples are very well distinguished from a gray mold of strawberry and tomato and from paprika, while widely overlapped in some cases, such as powdery mildew. We evaluated the crop disease prediction model using the ROC curve (Fig. 2b). The risk score was used as a rank to classify the crop's status and draw the ROC curve. The average of AUROC was 0.917, indicating that the model could classify the pests with high performance. In particular, the classification performance is consistent with the overlapping of risk scores. It is shown that the classification performance came out low in the case of powdery mildew because the risk scores of normal and infested samples overlap a lot. The low prediction performance and high overlap of risk scores can be interpreted that there are other influential factors in addition to the acquired environmental variables affecting the outbreak of the pest infestation. From the data science perspective, the model finds the relationships between the changes in growth environments and the outbreak of pests and diseases on the linear axis between 0 and 1. It learns how environmental differences lead to differences in risk scores. The overlaps occur if it is hard to explain the outbreak with only the environmental factors we used. Environmental variables such as solar radiation and relationships between host and fungi can be considered in further to improve the performance of the model [32, 33]. Experimental settings can be adequately adjusted for individual crops, pests, and diseases to alleviate the overlaps. Settings like two days of time window used in this study may be insufficient to train the onset of some rapid pests like powdery mildew [34, 35]. This problem can be

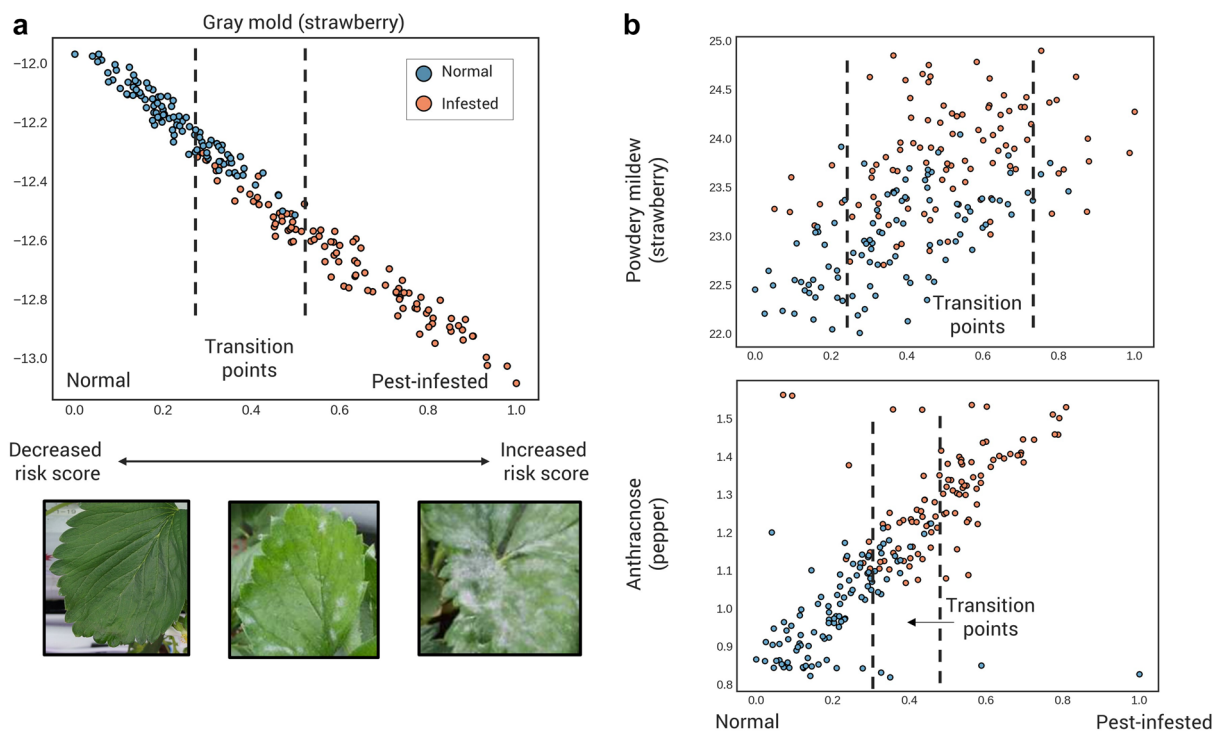


Fig. 1 Two-dimensional embeddings of test data. Two-dimensional representations of test data and assumed transition stages of **a** strawberry gray mold data, **b** strawberry powdery mildew data, and **c** pepper anthracnose data. And **a** examples of strawberry leaf images of data located in normal, transition point, and pest-infested each

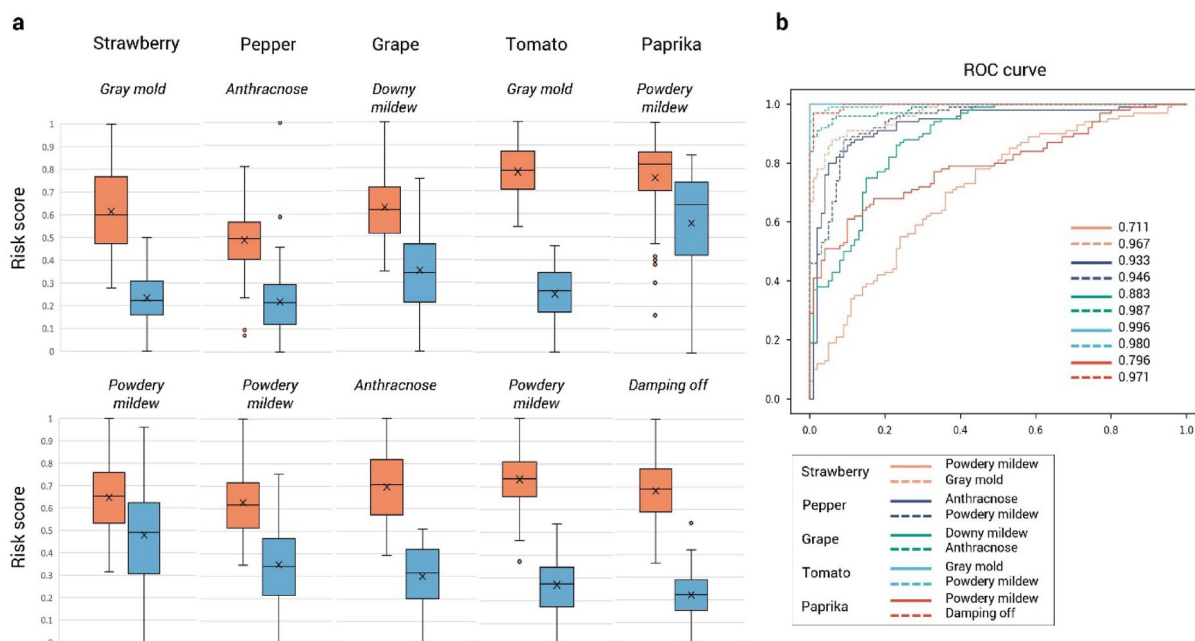


Fig. 2 Risk score distributions and prediction result. **a** Risk score distributions of normal and infested samples per crop and pest. **b** ROC curve and AUROC of prediction result for every crops and pests. The color of the curve represents the type of crop and the shape of the line represents the type of pests for each crop

alleviated by closely monitoring the growth environment or adjusting the number of sampling times while generating the training data.

We further investigated the relationship between disease severity and risk score. Using AI-hub's paprika damping off data, which includes the disease severity index, the predicted risk score is shown for each observation (Fig. 4a). The model is trained by a binary label, but it is shown that the risk score tends to increase as the severity of the disease increases. It shows that the prediction model can find a relationship between the exposed environment of crops and the occurrence and development of pests and diseases. However, since the difference in risk scores between infestation stages is small compared to the difference between normal and infested samples, it is crucial to maintain crops in a normal state by controlling the facility's environment before the transition point. Unlike binary classifiers, the prediction model uses risk scores as a surrogate measure for infestation prediction and can help tactical decision-making for pest prevention and management as a descriptive model. In order to test the possibility of a decision-making assistant for crop growth environment management, changes in the latent space were observed while adjusting the environmental data for one of the test data of strawberry gray mold and powdery mildew in the plant disease-causing dataset (Additional file 1: Fig. S3). The model can be used to prevent pests and diseases by managing the environment to the direction of maintaining a normal state by lowering the risk score caused by the movement of data point in the latent space led by the change in the growth environment.

Also, we tested our model with the in-house dataset, which contains strawberry gray mold data, to show the model trained with the plant disease-causing dataset can be applied to other datasets. The test data was embedded in the latent space and showed that the normal and infested samples were well aligned (Fig. 3b). The risk score distributions of the normal and infested samples were also consistent with the test results of the plant disease-causing dataset. Also, the prediction performance was evaluated with AUROC, and shows decent performance with AUROC of 0.8755 (Fig. 3c, d). However, the range of predicted risk scores was different. It may be caused by the difference in species of strawberry or fungi and environmental variety due to the facility, such as soil quality. This means some degree of overfitting occurs because input data cannot include every influential variable for the pest. Alignment corresponding to the risk of pest infestation is possible even with the pre-trained model, but the scale of the risk score is hardly matched. It shows that the fine-tuning stage is essential before applying the model to the individual facility.

Discussion

In this paper, we created a deep learning model to predict the presence or absence of pests using crop growth environmental data. It was shown that it is possible to control the environment for early diagnosis and prevention of pests and diseases through continuous environmental data monitoring. To this end, public and in-house data regarding crops' state and growth environment were collected. Moreover, to deal with large-scale heterogeneous data, we built an independent model for variables such as types of environmental variables, measurement intervals, and types of crops. Through this, it is possible to predict the status of a crop from its environmental data of the past two days. At the same time, the risk score can be extracted through the latent space. Unlike the binary classification, the risk score revealed a transition stage connecting the normal and disease groups. Early diagnosis and prediction before the outbreak will be possible during the transition stage. If a crop is located in the transition stage before the occurrence of pests and diseases, it is possible to prevent them by controlling the environment to reduce the risk score. This paper showed that environmental control is possible by reducing the risk score and moving the data close to the normal state by observing the pattern in the latent space that appears while changing the environmental variables from the point of interest through simulation.

Still, some challenges must be overcome to apply this model to existing facilities or farms. Many factors affect crop pests and diseases compared to the environmental variables measured in this study. There are plenty of factors that are not included in this study but are known to influence the growth of crops, such as soil components, soil pH, microbial properties, precipitation, and air current. In addition, in open-farm conditions, some variables are hard to quantify for data collection such as weather events, pollinators, and weeds [36, 37]. For example, compared to indoor farms that can artificially control many environmental variables, open-field farms have many variables, so it is difficult to train a model considering all of them. It is highly required to develop a multimodal monitoring system that collects high-resolution environmental data by quantifying individual features, as various environmental factors listed above are known to affect crop growth, diseases, and pests. In addition, considering relationships between different diseases or pests and optimal crop growth conditions, the model needs to be more complex. Resilience and resistance against a harsh environment and optimal growth environment are different, corresponding to the growth stage of crops [38]. Considering this, we can expect an increase in performance and practicality by adding temporal factors or the acquired growth stage data of target crops in

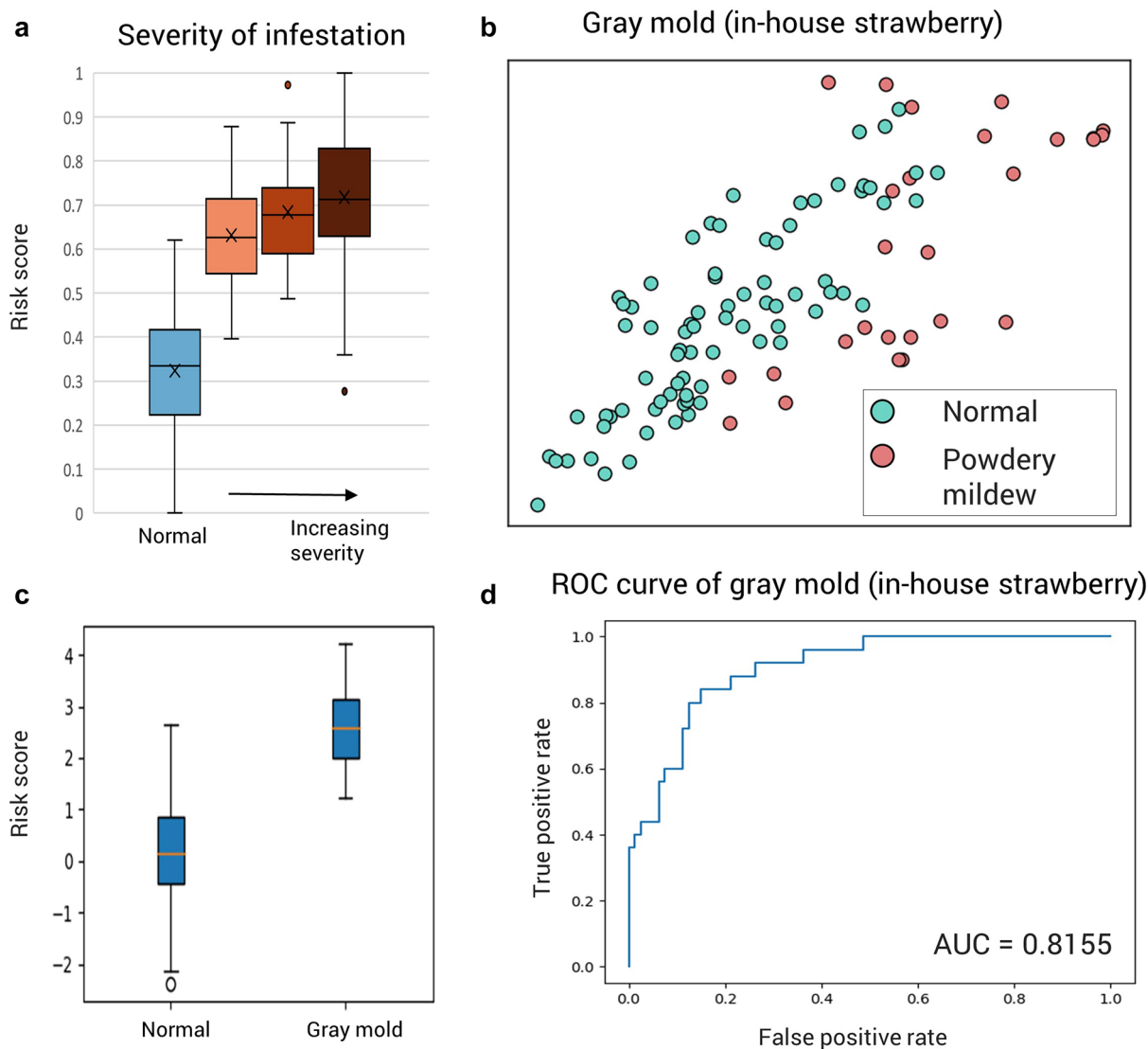


Fig. 3 Validation of the trained model. **a** Risk score distributions corresponding to the severity of observed infestation. Severity of paprika damping off increases from orange plot to brown plot. **b** Validation of gray mold prediction model using independent in-house dataset (latent space). **c** Risk score distributions of in-house samples; gray mold and normal. **d** Model evaluation of in-house gray mold prediction

the aspect of the life cycle to the training phase. The construction of a database suitable for these tasks must be preceded.

However, this model is highly versatile as it can be optimized and applied in various situations by adjusting parameters such as the type of environmental variable or the period of environmental recording. The resolution, performance, and usability of the model can be increased with additional sensors on various environmental variables. In addition to environmental data, various time-series information that can be acquired by sensors can be utilized in this method. In an experimental environment where environmental variables can be finely controlled,

this model can be used for physiological studies such as crop-pest interactions and crop reactions to environmental changes. In addition, in well-established facility farms, early management or prevention of pests and diseases can help improve crop quality, increase yield, and reduce required labor.

Methods

A large-scale dataset containing data on whether crops were infested with pests and diseases while growing and environmental information recorded during that period is needed to create a model that predicts the condition of crops from their growing environments. We used

large-scale public database, ‘plant disease-causing dataset’ for the data, reduced the dimensionality of the data, and built a model to predict the condition of crops using deep-learning methodology.

Data acquisition

Large-scale data were collected to build a model for predicting the occurrence and risk of pests and diseases with crop growth environment data. A public database named ‘plant disease-causing dataset’ provided by AI-hub is used. The dataset is data on disease and disease according to the environment of crops grown in several facilities. It provides disease outbreak data and cumulative environment data for at least 48 h before the observation time (Fig. 1a). Environment data includes at least temperature and humidity data and additionally collects CO₂, solar radiation, dewpoint, etc., depending on the facility equipment. We acquired environment and pest infestation data from the database for five crops and two pest infestations per crop. The number of data used in this study is shown in Table 1. The collected data includes various environmental data, temperature and humidity for all data, and other environmental data depending on the facility. With temperature and humidity recordings, grape data contains dewpoint data, and strawberry, pepper, and tomato data contains dewpoint and CO₂ concentration, while paprika data contains solar radiation measures. The diversity of data is a problem that inevitably arises due to the environmental information collection facilities of actual farms or the diversity of crop types.

In addition, as in-house data for validating the model trained using AI-hub, data on strawberry gray mold measured from November 2020 to May 2021 were collected. During the period, air temperature, solar radiation, relative humidity, and dewpoint of the strawberry facility were collected as environmental information every 10 min, and the percentage of infested strawberries

in the bed was measured. Unlike the plant disease-causing dataset that contains binary crop condition information, the in-house data is obtained as a continuous value because the percentage of diseased crops out of the total crops is measured.

Prediction model

A deep learning-based methodology was applied to build a model that predicts crop conditions at a specific point in time through past growing environments [39]. The autoencoder-based model was developed to predict the condition of crops with environmental data measured for the past two days from the forecasting point (Fig. 4b). The model used in this study takes an input vector representing two days of observations. Two days of cumulative environmental data were converted into a one-dimensional vector to make it an input for the model. For all environmental variables included in the data, the average, minimum, and maximum values of the measured values during the period are calculated, and all of these are concatenated to form a one-dimensional vector representing that period of observation. As a result, for a facility measuring n environmental variables, an input vector of length $3n$ is created (Fig. 4a).

The prediction model consists of an encoder composed of two layers with the same length as the input vector, a decoder having the same structure as the encoder, and a two-dimensional latent vector. The model was developed in a PyTorch environment based on modified loss functions suggested by the original paper [39]. Ordinary autoencoders are trained using the error between the input vector and the reconstructed vector as loss, but in this study, how well it predicts the label representing the state of the crop is added to the loss function (Additional file 1: Fig. S2). In the training phase, the mean squared error between the input and reconstructed vectors and the distance between the reference point on the alpha axis corresponding to the label was added and used as a loss function to make the normal and pest data points agglomerate respectfully. We set the reference point $\alpha=0$ for normal data and $\alpha=1$ for pest data. Since the data points are aligned along the alpha axis, we get the continuous value of the alpha coordinate of the data point, which can be considered as a risk score. So the prediction model can be used as an autoencoder-based regression model, although we trained the model using binary-labeled data. We trained a model in AI hub data to classify environmental vectors based on labels such as the presence or absence of crop diseases or pests. In all cases, we used 80% of the data as training data and the rest as test data. In the learning process, the learning rate, weight decay, and epoch were optimized for crop and pest respectively, while batch size is fixed to 64. After

Table 1 Data statistics acquired from the plant disease-causing dataset

| Crop | Disease | # of data |
|------------|--|-----------|
| Strawberry | Gray mold (<i>Botrytis cinerea</i>) | 16,000 |
| | Powdery mildew (<i>Spherotheca humuli</i>) | 10,000 |
| Pepper | Anthracoise (<i>Colletotrichum acutatum</i>) | 13,000 |
| | Powdery mildew (<i>Leveillula taurica</i>) | 12,000 |
| Grape | Downy mildew (<i>Plasmopara viticola</i>) | 8500 |
| | Anthracoise (<i>Colletotrichum fructicola</i>) | 2500 |
| Tomato | Gray mold (<i>Botrytis solan</i>) | 12,000 |
| | Powdery mildew | 10,000 |
| Paprika | Powdery mildew (<i>Leveillula taurica</i>) | 26,000 |
| | Damping off (<i>Rhizoctonia solini</i>) | 1000 |

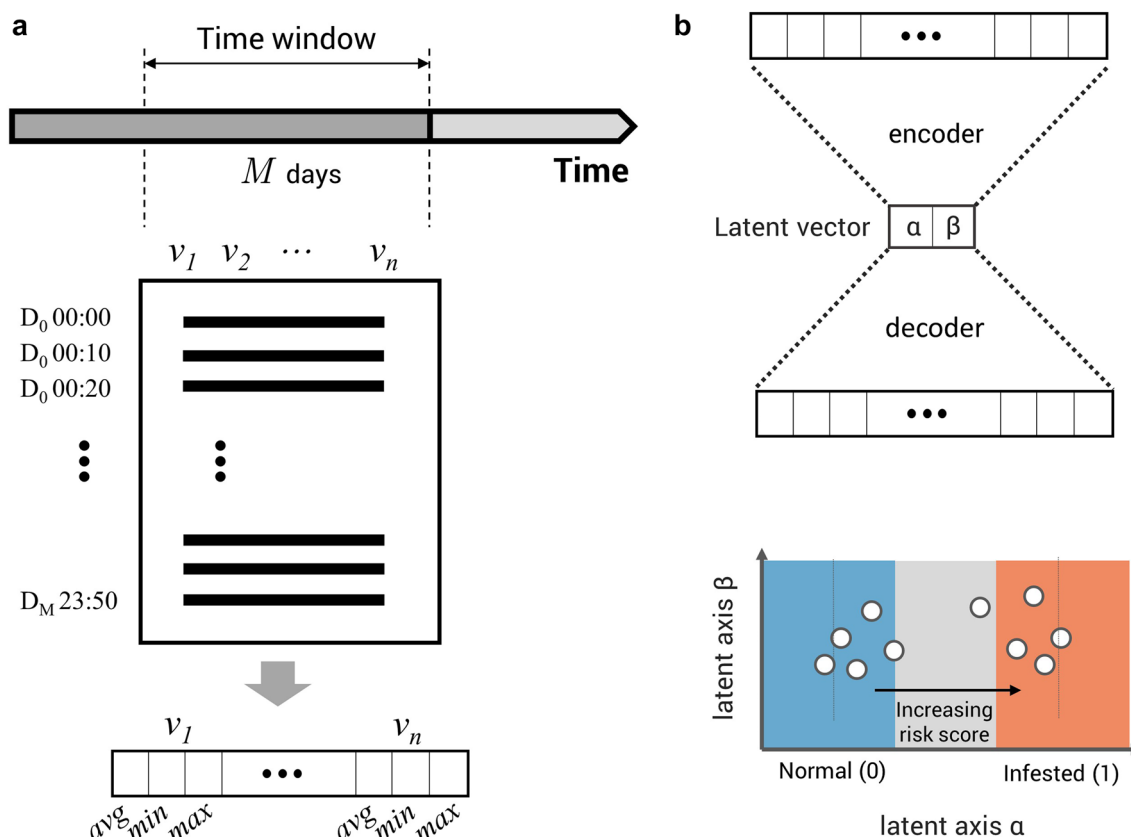


Fig. 4 A schematic of data acquisition and the prediction model. **a** A simplified representation of the format of data and transformation to one-dimensional vector obtained during the data measurement period. M is set to 2 as we acquired recordings of 48 h. **b** schematic of autoencoder structure and optimal embeddings to the latent space

the hyperparameter optimization, five-fold cross validation was performed. The risk scores from each trial was used for further analysis.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-023-01122-x>.

Additional file 1: Figure S1. Embedded points of test data on two-dimensional latent space. **Figure S2.** Change of loss during the training. **Figure S3.** Movement of data points with simulation.

Author contributions

The authors confirm their contribution to the paper: study conception and design: SL, CMY; data collection: SL; analysis and interpretation of results: SL, CMY; draft manuscript preparation: SL and CMY. All authors reviewed the results and approved the final version of the manuscript.

Funding

This work was carried out with the support of "Agriculture Technology Management Research Program for Agriculture Science & Technology Development (Project No. PJ01533203)" Rural Development Administration, Republic of Korea.

Data availability

The dataset used in this study is a public dataset with limited access that can be used after approval by the National Information Society Agency (NIA). Details of the dataset can be found on the AI-Hub website. (<https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=525>) The in-house data of this study are available from the corresponding author, C. M. Yun, upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

Author C. M. Yun is the CEO of Sherpa Space Inc. Remaining authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 September 2023 Accepted: 4 December 2023

Published online: 14 December 2023

References

- Hardwick NV. Disease forecasting. In: Jones DG, editor. *The epidemiology of plant diseases*. Dordrecht: Springer; 1998. (10.1007/978-94-017-3302-1_10).
- Savary S, Willocquet L, Pethybridge SJ, et al. The global burden of pathogens and pests on major food crops. *Nat Ecol Evol*. 2019;3:430–9. <https://doi.org/10.1038/s41559-018-0793-y>.
- Savary S, Bregaglio S, Willocquet L, et al. Crop health and its global impacts on the components of food security. *Food Sec*. 2017;9:311–27. <https://doi.org/10.1007/s12571-017-0659-1>.
- Sharma S, Kooner R, Arora R. Insect pests and crop losses. In: Arora R, Sandhu S, editors. *Breeding insect resistant crops for sustainable agriculture*. Singapore: Springer; 2017. (10.1007/978-981-10-6056-4_2).
- Mahmud MS, Zaman QU, Esau TJ, Price GW, Prithiviraj B. Development of an artificial cloud lighting condition system using machine vision for strawberry powdery mildew disease detection. *Comput Electron Agric*. 2019;158:219–25.
- Liu B. Sustainable strawberry production and management including control of strawberry powdery mildew. 2017. PhD Thesis. University of Hertfordshire.
- Yong D, Li Y, Gong K, Yu Y, Zhao S, Duan Q, Ren C, Li A, Fu J, Ni J, Zhang Y, Li R. Biocontrol of strawberry gray mold caused by *Botrytis cinerea* with the termite associated *Streptomyces* sp. sdu1201 and actinomycin D. *Front Microbiol*. 2022;13:1051730. <https://doi.org/10.3389/fmicb.2022.1051730>.
- Varah A, Ahodo K, Coutts SR, et al. The costs of human-induced evolution in an agricultural system. *Nat Sustain*. 2020;3:63–71. <https://doi.org/10.1038/s41893-019-0450-8>.
- Khan BA, et al. Pesticides: impacts on agriculture productivity, environment, and management strategies. In: Aftab T, editor, et al., *Emerging contaminants and plants. Emerging contaminants and associated treatment technologies*. Cham: Springer; 2023. (10.1007/978-3-031-22269-6_5).
- van Maanen A, Xu X-M. Modeling plant disease epidemics. *Eur J Plant Pathol*. 2003;109:669–82.
- Rossi V, Giosue S, Bugiani R. A-scab (Apple-scab), a simulation model for estimating risk of Venturia inaequalis primary infections. *OEPP/EPPO Bull*. 2007;37:300–8.
- Nath RK, Begum KH, Choudhury MR. Plant disease forecasting models. *Ind J Pure App Biosci*. 2020;8(4):454–61. <https://doi.org/10.18782/2582-2845.8280>.
- Wermelinger B, Oertli JJ, Baumgartner J. Environmental factors affecting the lifetables of *Tetranychus urticae* (Acari: Tetranychidae). III. Host-plant nutrition. *Exp Appl Acarol*. 1991;12:259–74.
- Kim JS, Jung C, Lee JH. Parameter estimation for temperature-dependent development model of *Tetranychus urticae* Koch: immature development. *J Asia-Pacific Entomol*. 2001;4(2):123–9. [https://doi.org/10.1016/S1226-8615\(08\)60113-2](https://doi.org/10.1016/S1226-8615(08)60113-2).
- Jung C, Kim D-S, Park Y-S, Lee J-H. Simulation modeling of twospotted spider mite population dynamics in apple and pear orchards in Korea. *J Asia-Pacific Entomol*. 2005;8(3):285–90. [https://doi.org/10.1016/S1226-8615\(08\)60247-2](https://doi.org/10.1016/S1226-8615(08)60247-2).
- Bregaglio S, et al. A public decision support system for the assessment of plant disease infection risk shared by Italian regions. *J Environ Manage*. 2022;317(1): 115365. <https://doi.org/10.1016/j.jenvman.2022.115365>.
- Edwards-Jones G. Knowledge-based systems for pest management: an applications-based review. *Pestic Sci*. 1992;36:143–53. <https://doi.org/10.1002/ps.2780360210>.
- Agatz A, Ashauer R, Sweeney P, Brown CD. A knowledge-based approach to designing control strategies for agricultural pests. *Agric Syst*. 2020. <https://doi.org/10.1016/j.agry.2020.102865>.
- Legler SE, Caffi T, Rossi V. A nonlinear model for temperature-dependent development of *Erysiphe nectar chasmothecia* on grapevine leaves. *Plant Pathol*. 2011. <https://doi.org/10.1111/j.1365-3059.2011.02498.x>.
- Technow F, Bürger A, Melchinger AE. Genomic prediction of northern corn leaf blight resistance in maize with combined or separated training sets for heterotic groups. *G3 (Bethesda)*. 2013;3(2):197–203. <https://doi.org/10.1534/g3.112.004630>.
- Huber DM, Haneklaus S. Managing nutrition to control plant disease. *Landbauforschung Volkenrode*. 2007;4(57):313–22.
- Lamichhane JR, Venturi V. Synergism between microbial pathogens in plant disease complexes: a growing trend. *Front Plant Sci*. 2015. <https://doi.org/10.3389/fpls.2015.00385>.
- Singh BP, Govindakrishnan PM, Ahmad I, Rawat S, Sharma S, Sreekumar J. INDO-BLIGHTCAST—a model for forecasting late blight across agroecologies. *Int J Pest Manag*. 2016. <https://doi.org/10.1080/09670874.2016.1210839>.
- Khatab A, Habib SE, Ismail H, Zayan S, Fahmy Y, Khairy MM. An IoT-based cognitive monitoring system for early plant disease forecast. *Comput Electron Agric*. 2019;166:105028. <https://doi.org/10.1016/j.compag.2019.105028>.
- Bhakta I, Phadikar S, Majumder K, et al. A novel plant disease prediction model based on thermal images using modified deep convolutional neural network. *Precision Agric*. 2023;24:23–39. <https://doi.org/10.1007/s11119-022-09927-x>.
- Newlands NK. Model-based forecasting of agricultural crop disease risk at the regional scale, integrating airborne inoculum, environmental, and satellite-based monitoring data. *Front Environ Sci*. 2018. <https://doi.org/10.3389/fenvs.2018.00063>.
- Fenu G, Maridina Mallocci F. An application of machine learning technique in forecasting crop disease. In: *Proceedings of the 3rd International Conference on Big Data Research (ICBDR '19)*. Association for Computing Machinery, New York, NY, USA, 76–82. 2020. <https://doi.org/10.1145/3372454.3372474>.
- Panigrahi KP, Das H, Sahoo AK, Moharana SC. Maize leaf disease detection and classification using machine learning algorithms. In: Das H, Pattnaik P, Rautaray S, Li KC, editors. *Progress in computing, analytics and networking. Advances in intelligent systems and computing*, vol. 1119. Singapore: Springer; 2020. (10.1007/978-981-15-2414-1_66).
- Xu Y, Mao Y, Li H, et al. A deep learning model for rapid classification of tea coal disease. *Plant Methods*. 2023;19:98. <https://doi.org/10.1186/s13007-023-01074-2>.
- Yang C, Baireddy S, Méline V, et al. Image-based plant wilting estimation. *Plant Methods*. 2023;19:52. <https://doi.org/10.1186/s13007-023-01026-w>.
- Delnevo G, Girau R, Ceccarini C, Prandi C. A deep learning and social IoT approach for plants disease prediction toward a sustainable agriculture. *IEEE IoT J*. 2022;9(10):7243–50. <https://doi.org/10.1109/JIOT.2021.3097379>.
- Amsalem L, Freeman S, Rav-David D, et al. Effect of climatic factors on powdery mildew caused by *Sphaerotheca macularis* f. sp. *Fragariae* on strawberry. *Eur J Plant Pathol*. 2006;114:283–92. <https://doi.org/10.1007/s10658-005-5804-6>.
- Pogoda M, Liu F, Douchkov D, Djamei A, Reif JC, Schweizer P, et al. Identification of novel genetic factors underlying the host-pathogen interaction between barley (*Hordeum vulgare* L.) and powdery mildew (*Blumeria graminis* f. sp. hordei). *PLoS ONE*. 2020;15(7):e0235565. <https://doi.org/10.1371/journal.pone.0235565>.
- Yuan L, Pu R, Zhang J, et al. Using high spatial resolution satellite imagery for mapping powdery mildew at a regional scale. *Precision Agric*. 2016;17:332–48. <https://doi.org/10.1007/s11119-015-9421-x>.
- Uloth MB, You MP, Barbetti MJ. Plant age and ambient temperature: significant drivers for powdery mildew (*Erysiphe cruciferarum*) epidemics on oilseed rape (*Brassica napus*). *Plant Pathol*. 2018;67(2):445–56. <https://doi.org/10.1111/ppa.12740>.
- Liliane TN, Charles MS. Factors affecting yield of crops. *Agronomy-climate change & food security*, 2020, 9.
- Lin Z, Wang S, Fu R, Ting KC, Lin T. Data-driven modeling for crop growth in plant factories. In: Ma S, Lin T, Mao E, Song Z, Ting KC, editors. *Sensing, data managing, and control technologies for agricultural systems*. Springer: Agriculture Automation and Control; 2022. (10.1007/978-3-031-03834-1_5).
- Dahal K, Li X-Q, Tai H, Creelman A, Bizimungu B. Improving potato stress tolerance and tuber yield under a climate change scenario—a current overview. *Front Plant Sci*. 2019. <https://doi.org/10.3389/fpls.2019.00563>.
- Lee S, Kim H, Lee D. Linearization autoencoder: an autoencoder-based regression model with latent space linearization. *BioRxiv*. 2022. <https://doi.org/10.1101/2022.06.06.494917>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.