Studentship Project: Annual Progress Report 10/2020 to 12/2022

Student Name:	**Jack Stevenson**	**AHDB Project Number:**	**SF/TF 170/a**
Project Title:	**Precision Agriculture: Using AI and Expert models to predict strawberry yields with respect to Powdery Mildew**		
Lead Partner:	University of Lincoln/NIAB EMR		
Supervisor:	Professor Stefanos Kollias, Dr. Mark Else		
Start Date:	24/09/2020	**End Date:**	24/09/2024

1. Project aims and objectives

Project Aims:

- Map variation of temperature and humidity using multiple observation signals, with the final goal to utilise 1-5 observation signals to map to different spatial resolutions.
- Use the developed variation model to estimate yield and Powdery Mildew Disease.

Current Objectives:

- Utilize different signals from the raw data to build a simple function of tunnel variation and fit that to a different tunnel design/setup.
- Use this estimated state to perform state forecasts for the different signals to the growing system
- Apply these state forecasts to models of disease and yield.

2. Key messages emerging from the project

It is likely that we can accurately estimate the future temperature and humidity of a polytunnel using Artificial Intelligence. Our current ANN models obtain a validation accuracy of approximately 30% using raw temperature data from polytunnels. These models are simplistic in nature compared to state-of-the-art models such as Numerical Weather Prediction Models [1] by having few layers to fully express the problem domain.

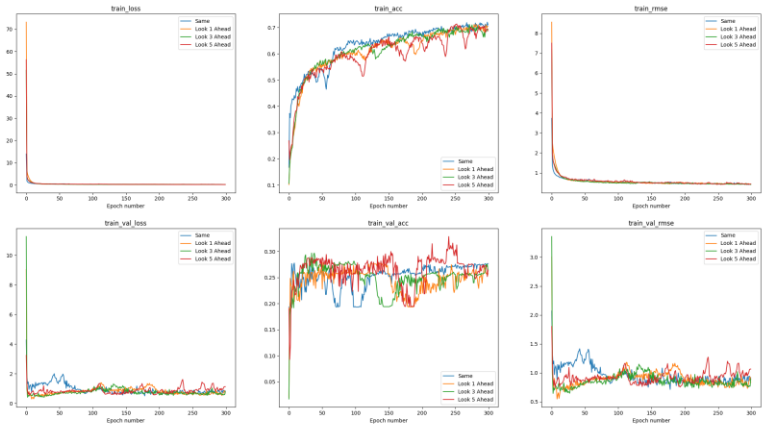
3. Summary of results from the reporting year

The data that has been collected this reporting year covers multiple spatial points in 2 separate polytunnels. Observation points recorded temperature and humidity values in 10-minute intervals. This provides a wealth of information that can be utilized for the models and allow for fine-grain optimisation during the working day given that some models, such as NIAB EMR's Powdery Mildew model having a daily resolution [2], PremonitionNet, which is a yield forecast system utilising 4-hour intervals for yield forecasting [3] and models of greenhouse environments for yield prediction [4].

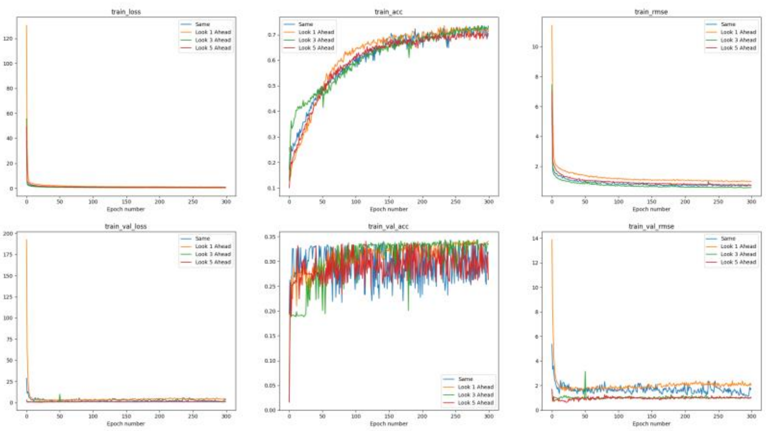
As mentioned above we have developed several models for analysing this data for both same step prediction and N step ahead prediction. Each model is based on an initial method layer, convolutions, recurrent and fully connected layers, followed by an output fully connected layer. We use a hyperband method of hyper-parameter optimisation to identify suitable model architecture parameters for each of the prediction problems.

The first prediction problem refers to same timestep prediction. This problem serves as an introduction to understanding the overall state of the tunnel at a given timestep based on observations for imputing data along other rows. Using an 80-20 train-test split, we observe a test accuracy of approximately 30% for each model over 300 epochs of data for each model type (seen as the blue lines on figures 1,2,3).

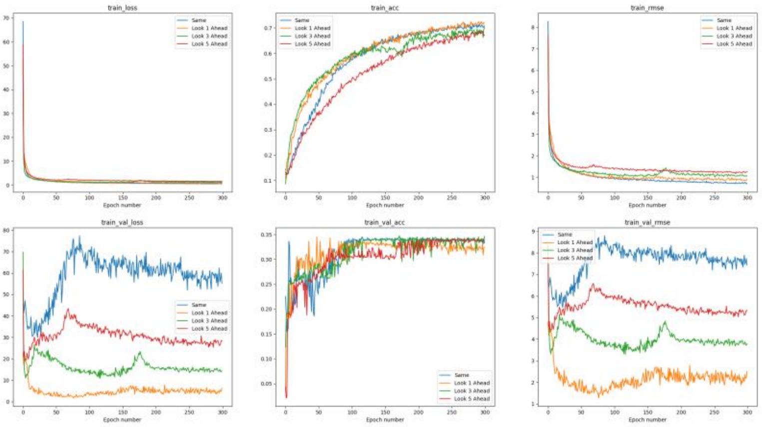
For the next N step prediction problem, we took the data that was used to train the same step models and offset the target data by N timesteps. This allows us to begin to learn the patterns for a future point in time, e.g., 50 minutes ahead based on the current observations. The Orange, green and red line in figures 1,2,3 indicate the training and validation results for 1 time step ahead, 3-time steps ahead and 5-time steps ahead respectively. From these graphs we can see that the architectures all plateaux around the 30-35% mark, suggesting that the current model architecture is unable to fully represent the problem domain. Our current approach uses a spatiotemporal model which accepts spatially arranged data over sliding windows at its input and predicts temperature at missing locations N timesteps ahead. The performance of the model is highly improved and forms the basis for extending it to both humidity and yield prediction, in a multitasking learning framework.



(Figure 1, MSE loss, Accuracy and RMSE for the convolutional networks on both the train and test datasets.)



(Figure 2, MSE loss, Accuracy and RMSE for the LSTM networks on both the train and test datasets.)



(Figure 3, MSE loss, Accuracy and RMSE for the dense networks on both the train and test datasets.)

4. Key issues to be addressed in the next year

During the data collection experiments being run the previous growing season, there were a few problems that occurred and will be discussed here. The first problem is the reliance on internet, in particular Wi-Fi, as Wi-Fi is needed to allow access to the data from across the country. This can be mitigated by utilising other technology, such as Ethernet over Power and 3G connectivity. These can work as fallback systems, however in the case of the polytunnel consideration, testing needs to be done to ensure the reliability of these technologies. The next major problem is with the aggregation of data from the sensors to the computer. Whilst we were theoretically able to connect to a device from 50m away in the tunnel, the connection was not stable enough to allow for the download process to occur, thus the data was never retrieved from devices automatically. There are 2 solutions to this problem, firstly, have a wider array of Bluetooth antenna that can be spread across the tunnels and use software to assign a list of sensors to each antenna. An alternative solution is to have more aggregation computers spread around the tunnel and assign each computer a list of sensors. The effectiveness of these approaches is dependent on the amount of interference the sensors will experience, so a higher and more central positioning of the aggregation computer would enable more data to be collected efficiently.

In terms of the data collected, the sensor downloads often had repeated datapoints, given that they were obtaining the history of observations for each sensor. In addition, timing issues due to unstable connections caused the downloaded data timestamps to be off by up to an hour. Both issues can be addressed in software by using the initial observation as a synchronisation point and using the known time of sensor activation as a base for setting the timestamps.

Furthermore, this season had passive data collection, whereas a more active data collection for useful targets such as yield and powdery mildew detection will be undertaken, where we will be working with farm managers to have a checklist for understanding the condition of the plants in a suitable timescale. By taking an active approach to the data collection process, we will be able to further improve the effectiveness of our models in the spatiotemporal prediction of yield and disease.

We will also be looking at deploying our developed model for live site reporting of the environmental state in hourly and daily intervals. This will serve as an understanding of the effectiveness of the model in a production setting, as well as begin to deconstruct the factors that lead up to its prediction.

5. Outputs relating to the project

(events, press articles, conference posters or presentations, scientific papers):

Output	**Detail**
Presentation at CTP	**Summary of the data gathering work and how it will be used**

6. Partners (if applicable)

Scientific partners	
Industry partners	
Government sponsor	

[1] Schultz, M.G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L.H., Mozaffari, A. and Stadtler, S., (2021) Can deep learning beat numerical weather prediction?. *Philosophical Transactions of the Royal Society A*, *379*(2194), p.20200097.
[2] Xu, X., Wedgewood, E., Berrie, A., Passey, T. and Hall, A., (2020) AHDB Project SF157: Improving integrated disease management in strawberry
[3] Onoufriou, G., Hanheide, M. and Leontidis, G., (2022) Premonition Net, A Multi-Timeline Transformer Network Architecture Towards Strawberry Tabletop Yield Forecasting. *arXiv preprint arXiv:2211.08177*.
[4] Gong, L., Yu, M., Jiang, S., Cutsuridis, V. and Pearson, S., 2021. Deep learning based prediction on greenhouse crop yield combined TCN and RNN. *Sensors*, *21*(13), p.4537.