Grower Summary

The Augmented Agronomist

SF/TF 170

Annual report 2021

Project title:	The Augmented Agronomist
Project number:	2155898, BB/S507453/1
Project leader:	George Onoufriou, University of Lincoln
Report:	Annual report, 2021-10
Previous report:	Annual report, 2020-10
Key staff:	Georgios Leontidis, University of Lincoln Marc Hanheide, University of Lincoln Mark Else, National Inst of Agricultural Botany
Location of project:	University of Lincoln Brayford Pool, Lincoln, LN6 7TS, England
Industry Representative:	Richard Harnden, Berry Gardens Growers Ltd, Unit 20 Wares Farm ME17 4BA
Date project commenced:	2018-11-30

DISCLAIMER

[The results and conclusions in this report are based on an investigation conducted over a one-year period. The conditions under which the experiments were carried out and the results have been reported in detail and with accuracy. However, because of the biological nature of the work it must be borne in mind that different circumstances and conditions could produce different results. Therefore, care must be taken with interpretation of the results, especially if they are used as the basis for commercial product recommendations.]

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

George Onoufriou

PhD Candidate

University of Lincoln

Signature .... Date ....

Report authorised by:

Georgios Leontidis

Associate Professor in Machine Learning

University of Aberdeen/ University of Lincoln

Signature .... Date ....

Marc Hanheide

Professor of Intelligent Robotics and Interactive Systems

University of Lincoln

Signature .... Date ....

Mark Else

Plant and Fruit Physiologist

National Institute of Agricultural Botany, East Malling Research

Signature .... Date ....

GROWER SUMMARY

Headline

Using machine learning/ artificial intelligence and (most likely) existing data sources describing environmental conditions like temperature and humidity, we can accurately and completely privately predict strawberry yield weeks in advance. This predictive performance ranges from 8% to 22% depending on the difficulty of the specific scenario being predicted for. This performance improves as more data becomes available which is necessary to be able to use more advanced neural networks. We produce these predictions without decrypting the input data at all and returning a merely transformed cyphertext privately.

Background

Inaccurately forecasting fruit yields like strawberries directly causes issues such as fruit waste, and insufficient production to meet contractual obligations. Fruit waste through underprediction leaves a surplus of fruit which then either must be sold at a reduced rate or destroyed. Over prediction leads to the opposite problem of having insufficient fruit to meet contractual obligations such as those agreed weeks in advance to retailers. This often means alternative supply must be sought often from abroad to cover this shortfall, since conditions that likely caused this shortfall are likely to affect geographically adjacent producers in the same country, further exasperating the additional costs. Currently yield forecasting varies wildly from grower to grower which in recent years has been, at worst ±50% of the actual yield.

Summary

Deep learning, a subset of machine learning and artificial intelligence is a state-of-the-art predictive technology where there is data available to train/ use it. Often the better data available the better the predictive performance becomes. However, data availability is often scarce. This project sets to do two things primarily; To show how deep learning can provide an invaluable accurate predictive service to both growers and agronomists. To lower barriers to data sharing to enable even the most privacy concerned growers to share data to gain a benefit from these state-of-the-art neural networks, using state-of-the-art quantum resistant cryptography like fully homomorphic encryption. This way we can provide accurate predictions using likely already existing data sources, improve on food waste and costs, and maintain complete privacy for the growers and their methods. We exemplify this on strawberries but fundamentally is agnostic of this specific application. If there is data that describes a relationship well, then a neural network can be taught to predict it, privately.

Financial Benefits

It is extremely difficult to quantify the specific financial benefits of such a predictive system without specific context with which to apply. However, in the strawberry domain we have been provided estimates at 26M/year losses across the UK industry due to the over and under-prediction of strawberry yields, even with agronomists who usually perform within ±17%.

The costs of improving this yield prediction, and thus reducing losses amounts to expertise, as the hardware and data necessary are likely to already exist. In the likely scenario that machine learning expertise does not exist in-house to operate these models, then external expertise is necessary. This could take the form of hiring a machine-learning engineer to setup on-site data processing. Or hiring a third party for continued deep learning as a service. However, currently only we are likely to be able to provide encrypted deep learning as a service due to its high barriers to entry if privacy is of primary concern.

Action Points

The primary action points to be able to exploit deep learning/ machine learning are the existence of data and expertise. Even if the intention is not to utilise this specific system, data is invaluable and should be recorded wherever possible for future use. The primary factors that relate to good usable data, are consistency, representativeness, and granularity. Collecting good consistent data usually means automation, to have sensors on site collecting environmental conditions such as temperature, light intensity, humidity continuously without/ as few gaps/ breaks as possible, over as long a period as possible, preferably years.

For data to be representative then it should be directly related to the outcomes being predicted, in the case of strawberries that usually means locality. The sensors should collect data as representative of what is experienced by the strawberries as possible. MET office data for example is usually too distant and not representative of the climate the berries have specifically experienced and been affected by. Granularity is effectively the descriptiveness of the data, having more than one sensor in multiple locations makes the data more granular to the climate conditions especially across larger sites. The regularity of the sensor taking readings also improves granularity as it helps describe the environment over time in more detail, giving machines more information with which to base their predictions on.